

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:11 ; Search time 51.35 Seconds
(without alignments)
64.892 Million cell updates/sec

Title: US-09-684-215A-17
Perfect score: 148
Sequence: 1 TAAADNFQLSGGGGAIPIGMAAINGQI 30

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	30	22	AAU69788
2	148	100.0	30	22	AAU01143
3	148	100.0	30	22	AAU699028
4	148	100.0	132	18	AAW32422
5	148	100.0	132	18	AAW32354
6	148	100.0	132	18	AAW32354
7	148	100.0	132	19	AAW81657
8	148	100.0	132	19	AAW64294
9	148	100.0	132	20	AAU39096
10	148	100.0	132	20	AAU38859
11	148	100.0	132	22	AAU69898
					AAU69906
					Mycobacterium tube

12	148	100.0	132	22	AAU01253	Mycobacterium tube
13	148	100.0	132	22	AAU01261	Mycobacterium tube
14	148	100.0	132	22	AAU01889	M. tuberculosis pa
15	148	100.0	224	22	AAU69899	Human prostate pro
16	148	100.0	224	22	AAU01254	Ra12-P510S-C const
17	148	100.0	231	20	AAU32071	Mycobacterium tube
18	148	100.0	304	22	AAU69902	Human /M. tubercul
19	148	100.0	304	22	AAU01257	Ra12-P775P-ORF3 co
20	148	100.0	355	18	AAW32435	Mycobacterium tube
21	148	100.0	355	19	AAW32367	Mycobacterium tube
22	148	100.0	355	18	AAW81670	M. tuberculosis im
23	148	100.0	355	19	AAW64307	Mycobacterium tube
24	148	100.0	355	20	AAU39109	M. tuberculosis an
25	148	100.0	355	20	AAU38972	M. tuberculosis re
26	148	100.0	355	20	AAU05000	Mycobacterium spec
27	148	100.0	355	22	AAU81110	Mycobacterium tube
28	148	100.0	355	22	AAU01890	M. tuberculosis an
29	148	100.0	379	20	AAU04830	Mycobacterium spec
30	148	100.0	400	22	AAU69907	Human prostate pro
31	148	100.0	400	22	AAU01262	Ra12-P510S-E2 cons
32	148	100.0	487	22	AAU63280	Chlamydia trachoma
33	148	100.0	518	22	AAU63276	Chlamydia trachoma
34	148	100.0	525	21	AAU13645	C. pneumoniae sero
35	148	100.0	525	22	AAU63213	Protein encoded by
36	148	100.0	543	22	AAU01905	M. tuberculosis an
37	148	100.0	583	22	AAU63281	Chlamydia trachoma
38	148	100.0	585	22	AAU63277	Chlamydia trachoma
39	148	100.0	619	22	AAU63270	Chlamydia trachoma
40	148	100.0	631	22	AAU63274	Chlamydia trachoma
41	148	100.0	646	22	AAU63272	Chlamydia trachoma
42	148	100.0	654	22	AAU63278	Chlamydia trachoma
43	148	100.0	683	22	AAU63282	Chlamydia trachoma
44	148	100.0	691	22	AAU63271	Chlamydia trachoma
45	148	100.0	700	22	AAU63279	Chlamydia trachoma

ALIGNMENTS

RESULT 1	AAU69788	standard; Peptide: 30 AA.
ID	AAU69788	
XX	AAU69788;	
AC	30-JAN-2002 (first entry)	
XX		
DE	M. tuberculosis antigen Ra12 amino acids 1-30.	
XX		
KW	Human; prostate cancer; ss; cytostatic; immunostimulant; tumour;	
KW	antigen; epitope.	
XX		
OS	Mycobacterium tuberculosis.	
XX		
PN	WO200173032-A2.	
XX		
PD	04-OCT-2001.	
XX		
PF	27-MAR-2001; 2001WO-US09919.	
XX		
PR	27-MAR-2000; 2000US-0536857.	
PR	09-MAY-2000; 2000US-0568100.	
PR	12-MAY-2000; 2000US-0570737.	
PR	13-JUN-2000; 2000US-0593793.	
PR	27-JUN-2000; 2000US-0605783.	
PR	10-AUG-2000; 2000US-0636215.	
PR	29-AUG-2000; 2000US-0651236.	
PR	06-SEP-2000; 2000US-0651279.	
PR	02-OCT-2000; 2000US-0679426.	
PR	10-OCT-2000; 2000US-0685166.	
XX		
XX	(CORI-) CORIXA CORP.	
PA		
XX		

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;
 XX WPI: 2001-639232/73.
 DR
 XX
 PT New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer -
 XX
 PS Example 17; Page 395; 579pp; English.
 CC
 CC The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen-presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumor protein, and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is a prostate specific
 CC peptide of the invention. The peptides either represent antigenic
 CC epitopes or domains of prostate specific proteins.
 CC
 SO Sequence 30 AA:
 Query Match 100.0%; Score 148; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. NO. 2.7e-15;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TAASDNFOLSGGGFAIPIGAMAIAGOI 30
 DB 1 taasdnfqlsgggfaipligamalaqgi 30
 RESULT 2
 AAM01143
 ID AAM01143 standard; Protein; 30 AA.
 XX
 AC AAM01143;
 XX
 DT 04-OCT-2001 (first entry)
 DE Mycobacterium tuberculosis antigen Ra12 first 30 amino acids.
 XX
 DE Mycobacterium tuberculosis antigen Ra12 first 30 amino acids.
 XX
 KM Human; prostate cancer; prostate-specific; diagnosis; vaccine;
 KM cytostatic; gene therapy; metastasis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200151633-A2.
 XX
 PD 19-JUL-2001.
 XX
 PE 16-JAN-2001; 2001WO-US01574.
 XX
 PR 14-JAN-2000; 2000US-0483672.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
 PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skelky YAW;
 PI Wang A, Heagner MJ;
 XX WPI: 2001-425873/45.
 DR
 XX New polynucleotide encoding a prostate-specific protein, for
 PT diagnosing, monitoring and treating prostate cancer in a patient and
 PT for use in vaccines -
 XX

PS Example 17; Page 394; 543pp; English.
 XX
 CC The present invention describes polynucleotide sequences (I) which encode
 CC prostate-specific proteins (II), (I) and (II) have cytostatic activity,
 CC and can be used in vaccine production and gene therapy. (I), (II),
 CC antibodies to (II), fusion proteins comprising (II), and isolated
 CC T cells prepared using (I) or (II) are used treat cancer in a patient.
 CC (I) and the antibodies are also used in the detection of cancer in a
 CC patient. The cancer that is diagnosed or treated is particularly
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
 CC (I) can be used for monitoring the progression of cancer in a patient.
 CC (I) and (II) can also be used to improve diagnostic and therapeutic
 CC methods for prostate cancer. They can indicate the level of metastasis
 CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
 CC AAM01318 represent polynucleotide and amino acid sequences used in the
 CC exemplification of the present invention.
 CC
 SO Sequence 30 AA:
 Query Match 100.0%; Score 148; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. NO. 2.7e-15;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TAASDNFOLSGGGFAIPIGAMAIAGOI 30
 DB 1 taasdnfqlsgggfaipligamalaqgi 30
 RESULT 3
 AAG99028
 ID AAG99028 standard; Protein; 30 AA.
 XX
 AC AAG99028;
 XX
 DT 25-SEP-2001 (first entry)
 DE Mycobacterium tuberculosis antigen Ra12 1st 30 amino acid sequence.
 XX
 DE Mycobacterium tuberculosis antigen Ra12 1st 30 amino acid sequence.
 XX
 KM Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
 KM chromosome 22q11.2; prostate-specific protein; chromosome 1;
 KM prostate specific antigen; PSA.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200134802-A2.
 XX
 PD 17-MAY-2001.
 XX
 PE 09-NOV-2000; 2000WO-US30904.
 XX
 PR 12-NOV-1999; 99US-0439313.
 XX
 PR 18-NOV-1999; 99US-0443686.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
 PI Kalos MD, Retter MW, Stolk JA, Day CH, Skelky YAW, Wang A;
 XX WPI: 2001-308785/32.
 DR
 XX Isolated polypeptide comprising at least an immunogenic portion of a
 PT prostate-specific protein, useful in the diagnosis and therapy of
 PT prostate cancer -
 XX
 PS Example 17; Page 294; 325pp; English.
 CC
 CC The present invention describes an isolated polypeptide (PI) comprising
 CC at least an immunogenic portion of a prostate-specific protein, or its
 CC variant. Also described are polynucleotides (NI) encoding (PI) and
 CC (NI) have cytostatic activity and can be used in vaccine production.
 CC The polypeptides, nucleic acids and antibodies from the present
 CC invention are useful in the diagnosis and therapy of prostate cancer.


```

AAM81657
ID AAM81657 standard; Protein; 132 AA.
XX
AC AAM81657;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide TbrA12.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PE 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR N-PSDB; AAV64450.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
PS Example 3; Page 97-98; 230pp; English.
XX
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC for inducing protective immunity against tuberculosis (TB). This sequence
CC can be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 132 AA;

```

Query Match 100.0%; Score 148; DB 19; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TAASDNFQLSOGGQGFAPIGQAMAIAGQI 30
   |||||
DB 1 taasdnfqlsggqgfaipligamataigqi 30

```

RESULT 7
AAM64294
ID AAM64294 standard; Protein; 132 AA.
XX
AC AAM64294;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbrA12.
XX
KW Tuberculosis; infection; diagnosis; antigen; TbrA12.
XX
OS Mycobacterium tuberculosis strain H37Ra.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.

```

XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR N-PSDB; AAV44342.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 3; Page 101; 250pp; English.
XX
CC This polypeptide comprises Mycobacterium tuberculosis soluble
CC antigen TbrA12. It is encoded by a DNA sequence (see AAV44342)
CC isolated from a M. tuberculosis strain H37Ra expression library
CC with rabbit anti-sera raised against M. tuberculosis supernatant.
CC The invention relates to compositions and methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAM64291-W64379)
CC comprising an antigenic portion of a soluble M. tuberculosis
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
CC as well as DNA sequences encoding such polypeptides, recombinant
CC expression vectors and transformed or transfected host cells. Also
CC claimed are methods and diagnostic kits for detecting M.
CC tuberculosis infection in a patient using the above polypeptides,
CC antibodies, or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 132 AA;

```

Query Match 100.0%; Score 148; DB 19; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TAASDNFQLSOGGQGFAPIGQAMAIAGQI 30
   |||||
DB 1 taasdnfqlsggqgfaipligamataigqi 30

```

RESULT 8
AAV39096
ID AAV39096 standard; Protein; 132 AA.
XX
AC AAV39096;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen TbrA12 amino acid sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PE 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX WPI: 1999-527409/44.
 DR N-PSDB; AA219040.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Example 3; Page 98; 299pp; English.
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's,
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
 CC AA219225 are used in the exemplification of the present invention.
 XX
 SQ Sequence 132 AA;

Query Match 100.0%; Score 148; DB 20; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1.4e-14;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGGFAIPIGOMAIAGOI 30
 ||||||||||||||||||||||||||||
 Db 1 taasdnfqlsgggfaiipigamalaqql 30

RESULT 9
 AA219083
 ID AA219083 standard; Protein; 132 AA.
 XX
 AC AA219083;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein Tbra12.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 XX vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX WPI: 1999-527416/44.
 DR N-PSDB; AA219040.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 PS Example 3; Page 138; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding

CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 132 AA;

Query Match 100.0%; Score 148; DB 20; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1.4e-14;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGGFAIPIGOMAIAGOI 30
 ||||||||||||||||||||||||||||
 Db 1 taasdnfqlsgggfaiipigamalaqql 30

RESULT 10
 AA219083
 ID AA219083 standard; Protein; 132 AA.
 XX
 AC AA219083;
 XX
 DT 30-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Ra12.
 XX
 KW Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;
 XX fusion protein; Ra12 antigen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200173032-A2.
 XX
 PD 04-OCT-2001.
 XX
 PF 27-MAR-2001; 2001WO-US09919.
 XX
 PR 27-MAR-2000; 2000US-0536857.
 PR 09-MAY-2000; 2000US-0568100.
 PR 12-MAY-2000; 2000US-0570737.
 PR 13-JUN-2000; 2000US-0593793.
 PR 27-JUN-2000; 2000US-0605783.
 PR 10-AUG-2000; 2000US-0636215.
 PR 29-AUG-2000; 2000US-0651236.
 PR 06-SEP-2000; 2000US-0679426.
 PR 02-OCT-2000; 2000US-0679426.
 PR 10-OCT-2000; 2000US-0685166.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Reiter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
 XX WPI: 2001-639232/73.
 DR
 XX
 PT New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer -
 XX
 PS Example 17; Page 531-532; 579pp; English.
 CC The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen-presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumour protein, and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the

XX WO200173032-A2.
PN
XX

PD 04-OCT-2001

PF 27-MAR-2001; 2001WO-US09919.
XX

PR	27-MAR-2000; 2000US-0536857.
PR	09-MAY-2000; 2000TS-0568100

PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793

PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215

PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.

PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.

XX
PA (CORI-) CORIXA CORP.

XX
PI Xu J, Dillon DC, Mitcham J

PI Fanger GK, Kelter MW, Scott
PI Li SX, Wang A, Skeiky YAW,

AA WPI; 2001-639232/73.
DR

XX
XX
XX

the diagnosis and treatment

PS Example 17; Page 533-534; 57

CC The invention relates to iso

CC antibodies raised against the

polypeptides. The antibodies

CC the antigen-presenting cells
CC T cells specific for a tumour

CC of cancer especially prostate
CC polynucleotide and/or polypeptide

CC response, and for treating c
CC detecting cancer. The presen

CC prostate specific polypeptide
XX
XX

sequence 224 AA;

Query Match	100.0%
100.0%	

Best Local Similarity 100.0%;
Matches 30; Conservative

QY 1 TAASDNFQLSQGGQGFAPIGQF

Db 8 taasdnfqlsgggqfaipigga

Search completed: August 13, 2002

100

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:32 ; Search time 12.82 Seconds

(without alignments)
57.158 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148
Sequence: 1 TAASDNFQLSGGGGFAIPICQAMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 30

Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%

Listing first 100 summaries

Database: Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCMTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	100.0	30	4	US-09-439-313-484 Sequence 484, App

ALIGNMENTS

RESULT 1
US-09-439-313-484
Sequence 484, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yugu
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439.313
CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo Saplen
US-09-439-313-484

Query Match 100.0%; Score 148; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSGGGGFAIPICQAMAIAGQI 30
|||||
Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQI 30

Search completed: August 13, 2002, 10:58:53
Job time: 21 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:12 ; Search time 25.01 Seconds
(without alignments) 115.261 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148

Sequence: 1 TAASDNFQLSGGSGFAIRPGAMAIAGQI 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_71:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	148	100.0	355	2	F70983	probable serine pr
2	104.5	70.6	361	2	S47170	hypothetical prote
3	98.5	66.6	354	2	A87242	probable secreted
4	56	37.8	296	2	F87383	hypothetical prote
5	55	37.2	464	2	C70821	probable serine pr
6	54	36.5	382	2	H86930	probable secreted
7	54	36.5	407	2	AG2150	serine proteinase
8	54	36.5	452	2	S75338	serine proteinase
9	54	36.5	452	2	T45448	probable serine pr
10	53.5	36.1	239	2	B54589	proteasome subunit
11	53	35.8	688	2	A12516	hypothetical prote
12	51.5	34.8	820	2	A86247	hypothetical prote
13	51	34.5	253	2	H90988	hypothetical prote
14	51	34.5	253	2	B85834	hypothetical prote
15	51	34.5	369	2	E86554	Fe-S oxidoreductas
16	51	34.5	369	2	G72069	conserved hypochet
17	51	34.5	429	2	AD1894	serine proteinase
18	50.5	34.1	1000	2	C82630	serine proteinase
19	50	33.8	362	2	T35287	probable secreted
20	49.5	33.4	358	2	A96739	hypothetical prote
21	49.5	33.4	416	2	S16306	membrane protein p
22	49.5	33.4	416	2	P90707	probable transport
23	49.5	33.4	416	2	B85558	probable transport
24	49.5	33.4	533	2	C83658	hypothetical prote
25	49.5	33.4	662	2	G89609	transketolase [imp
26	49	33.1	253	2	G64573	hypothetical prote
27	49	33.1	340	2	C96753	transcription fact
28	49	33.1	394	2	S74643	proteinase hboA (E
29	49	33.1	591	2	F83472	probable glycosyl

30	49	33.1	665	2	AG1117	transketolase homo
31	49	33.1	665	2	A11477	transketolase homo
32	49	33.1	1194	2	D49851	magnesium-protopor
33	48	32.4	350	2	S34557	hypothetical prote
34	48	32.4	506	2	G86806	lipopolysaccharide
35	48	32.4	697	2	T34704	probable transfera
36	48	32.4	822	2	E75523	AMP-dependent heli
37	47.5	32.1	341	2	S33464	hypothetical prote
38	47.5	32.1	342	2	A96581	hypothetical prote
39	47.5	32.1	581	2	T51580	ovule development
40	47.5	32.1	664	2	A86296	hypothetical prote
41	47.5	32.1	700	2	C86296	hypothetical prote
42	47.5	32.1	1487	2	G96827	protein F20817.10
43	47	31.8	257	2	A72395	oxidoreductase, sh
44	47	31.8	351	2	S29979	hype protein - Alc
45	47	31.8	412	2	JC1116	type III site-spec

ALIGNMENTS

RESULT 1
F70983
probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70983
R: Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; MUID:98295987
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-355 <CO>
A: Cross-references: GB:296071; GB:A123456; NID:93242254; PIDN:CAB09453.1; PID:g21819
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: pepA
C: Superfamily: Escherichia coli trypsin-like proteinase

Query Match 100.0%; Score 148; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. NO. 1.2e-13;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAASDNFQLSGGSGFAIRPGAMAIAGQI 30
DB 224 TAASDNFQLSGGSGFAIRPGAMAIAGQI 253

RESULT 2
S47170
hypothetical protein 34k - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999
C:Accession: S47170
R: Camerton, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
submitted to the EMBL Data Library, June 1993
A: Description: Isolation and characterisation of a 34kDa protein of Mycobacterium par
A: Reference number: S47170
A: Accession: S47170
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-361 <CAM>
A: Cross-references: EMBL:223092; NID:9505550; PIDN:CAA80638.1; PID:9505551
C: Superfamily: proteinase hboB

Query Match 70.6%; Score 104.5; DB 2; Length 361;

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW74457.1; PID:g17131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2758

Query Match 36.5%; Score 54; DB 2; Length 407;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 13 GGGFAIRIGQAMAIAGOI 30
| | | | | : | | | | :
Db 275 GIGFAIRIDAKAKAIATPL 292

RESULT 8
S77538
serine proteinase (EC 3.4.21.-) htra - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1204
C:Species: Synechocystis sp.

A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Matsuura, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.

A:Reference number: S74322; MUID:97061201
A:Accession: S77538
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <KAN>
A:Cross-references: EMBL:D90905; GB:AB01339; NID:g1652360; PIDN:BAAL17385.1; PID:g165248
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: htra
C:Superfamily: proteinase htrb
C:Keywords: hydrolase; serine proteinase

Query Match 36.5%; Score 54; DB 2; Length 452;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 15; Conservative 2; Mismatches 9; Indels 4; Gaps 1;

QY 1 TAAADNFOLSGGGFAIRIGQAMAIAGOI 30
| | | | | : | | | | :
Db 312 TAIIONAQ----GIGFAIRINKAQAIAOQL 337

RESULT 9
T45448
probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000
C:Accession: T45448
A:Reference number: 222967
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <JAM>
A:Cross-references: EMBL:AL035500; PIDN:CAB36690.1
A:Experimental source: cosmid U373
C:Genetics:
A:Note: MLCB373.28
C:Keywords: hydrolase; serine proteinase
F:182,224,305/active site: His, Asp, Ser #status predicted

Query Match 36.5%; Score 54; DB 2; Length 452;
Best Local Similarity 44.8%; Pred. No. 5.6;
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 2 AASDNFOLSGGGFAIRIGQAMAIAGOI 30
| | | | | : | | | | :
Db 328 ADSGDAQSGSIGLGFALPVDQAKRIADQL 356

RESULT 10

B54589
Proteasome subunit Y - human

C:Species: Homo sapiens (man)
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 03-May-1996
C:Accession: B54589
R:Akiyama, K.; Yokota, K.; Kagawa, S.; Shimbara, N.; Tamura, T.; Akioka, H.; Nothwang
Science 265, 1231-1234, 1994
A:Title: cDNA cloning and interferon gamma down-regulation of proteasomal subunits X
A:Reference number: A54589; MUID:94345396
A:Accession: B54589
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA

A:Residues: 1-239 <AKT>
A:Cross-references: GB:D29012
A:Note: at position 21, the translation given, shows an Ala
C:Superfamily: multicatalytic endopeptidase complex chain C5

Query Match 36.1%; Score 53.5; DB 2; Length 239;
Best Local Similarity 43.5%; Pred. No. 3.4;
Matches 10; Conservative 6; Mismatches 2; Indels 5; Gaps 1;

QY 11 GGGGFAIRIG-----QAMAIAG 28
: | | | : | | : | | | :
Db 141 ECGGGRSVPMGGMVRSFALIG 163

RESULT 11

AI2516
hypothetical protein all17313 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7

C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AI2516
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matsuura, A.; Iriju
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-688 <KUR>
A:Cross-references: GB:BA000020; PIDN:BAW77071.1; PID:g17134512; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all17313
A:Genome: plasmid

Query Match 35.8%; Score 53; DB 2; Length 688;
Best Local Similarity 42.1%; Pred. No. 12;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 10 SOGGGFAIRIGQAMAIAG 28
| : | | | : | | : | | :
Db 191 SEGGGLRVPPLGKTLALVG 209

RESULT 12
A86247
hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: AB6247
 R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719
 A:Accession: AB6247
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-820 <STO>
 A:Cross-references: GB:AB005172; NID:g5734727; PIDN:AAD4992.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein

Query Match 34.8%; Score 51.5; DB 2; Length 820;
 Best Local Similarity 45.2%; Pred. No. 24;
 Matches 14; Conservative 4; Mismatches 6; Indels 7; Gaps 2;

OY 3 ASDNQLS----QGQGFPIPIGQAMATAGQ 29
 Db 505 ATDNFSLSNKLGQGGFG--PYKGMLEGG 532

RESULT 13
 H90988
 hypothetical protein ECS2880 [imported] - Escherichia coli (strain O157:H7, substrain R1
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: H90988
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawaga, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: H90988
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <HAV>
 A:Cross-references: GB:BA000007; PIDN:BA836303.1; PID:g13362349; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: ECS2880
 C:Superfamily: Escherichia coli hypothetical protein b2072

Query Match 34.5%; Score 51; DB 2; Length 253;
 Best Local Similarity 47.6%; Pred. No. 8.2;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 9 LSQGGGFAIPIGQAMATAGQ 29
 Db 51 VSQGGGAMLAINEAMATVSQ 71

RESULT 14
 B85834
 hypothetical protein Z3240 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B85834
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 525-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: B85834
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <STO>
 A:Cross-references: GB:AB005174; NID:g12516274; PIDN:AAG57134.1; GSPDB:GN00145; UNCP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z3240
 C:Superfamily: Escherichia coli hypothetical protein b2072

Query Match 34.5%; Score 51; DB 2; Length 253;
 Best Local Similarity 47.6%; Pred. No. 8.2;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 9 LSQGGGFAIPIGQAMATAGQ 29
 Db 51 VSQGGGAMLAINEAMATVSQ 71

RESULT 15
 E86554
 Fe-S oxidoreductase [imported] - Chlamydia pneumoniae (strain J138)
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: E86554
 R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
 Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: AB6491; MUID:20330349
 A:Accession: E86554
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-369 <STO>
 A:Cross-references: GB:BA000008; NID:g8978883; PIDN:BA98719.1; GSPDB:GN00142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: CPj0513
 C:Superfamily: hypothetical protein AF0390

Query Match 34.5%; Score 51; DB 2; Length 369;
 Best Local Similarity 50.0%; Pred. No. 12;
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 8 QLSQGGGFAIPIGQAMAIA 27
 Db 272 RLNRSGGHAIPLEKSLMAYA 291

Search completed: August 13, 2002, 10:36:14
 Job time: 62 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:36:17 ; Search time 15.81 Seconds
(without alignments)
73.472 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148
Sequence: 1 TAAADFOLSGGCGFAPIGOMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5.

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.5	36.1	239	1 PSB6_HUMAN	P28072 homo sapien
2	49.5	33.4	416	1 YADA_ECOLI	P24077 escherichia
3	49	33.1	253	1 YBCK_ECOLI	P76395 escherichia
4	49	33.1	1194	1 BCHH_RHOC	P26162 rhodobacter
5	48	32.4	350	1 YCXA_EUGR	P31561 euglena gra
6	47.5	32.1	694	1 ABC7_MOUSE	Q61102 mus musculi
7	47	31.8	351	1 HYPE_ALCEU	P31905 alcaigenes
8	47	31.8	412	1 T3RE_BACCE	P25411 bacillus ce
9	46.5	31.4	225	1 Y130_MYCTU	O50603 mycobacteri
10	46.5	31.4	752	1 ABC7_HUMAN	O75027 homo sapien
11	46.5	31.4	928	1 PMP9_CHLPN	O92398 chlamydia p
12	46	31.1	1609	1 CYPE_MYCLE	O53114 mycobacteri
13	45.5	30.7	376	1 CARA_MYCTU	P71811 mycobacteri
14	45	30.4	396	1 F0EM_SHIDY	Q03583 shigella dy
15	45	30.4	541	1 601M_HAEIN	P44773 haemophilus
16	44.5	30.1	237	1 PSB6_RAT	P28073 rattus norv
17	44.5	30.1	238	1 PSB6_MOUSE	O60692 mus musculi
18	44	29.7	288	1 Y012_MYCPN	P75097 mycoplasma
19	44	29.7	351	1 MRAY_BORBU	Q44776 borrelia bu
20	44	29.7	382	1 AZAB_DIDMA	O77715 dieliphis m
21	43.5	29.4	119	1 CHAI_ANTPO	P02446 antherea p
22	43.5	29.4	776	1 VP4_ROTTH6	P26451 human rota
23	43.5	29.4	776	1 VP4_ROTTH6	P12473 rhesus rota
24	43	29.1	233	1 R0C1_SPTOL	P28644 spinacia ol
25	43	29.1	306	1 RBSK_HAEIN	P44331 haemophilus
26	43	29.1	432	1 Y1S4_YEAST	P40564 saccharomyc
27	43	29.1	444	1 PSF1_YEAST	Q12355 saccharomyc
28	43	29.1	461	1 SG2N_RAT	O56405 rattus norv
29	43	29.1	628	1 SNXI_HUMAN	O964f0 homo sapien
30	43	29.1	700	1 FLHA_CAUCR	O03845 caudobacter
31	43	29.1	1193	1 BCHH_RHOSH	Q947d5 rhodobacter
32	43	29.1	1362	1 PMD1_SCHPO	P36619 schizosacch
33	43	29.1	1530	1 RPL1_TRYBB	P08968 trypanosoma

34	42.5	28.7	110	1 RPO2_MYCLE	Q09c96 mycobacteri
35	42.5	28.7	617	1 PYS1_PSEAE	Q06583 pseudomonas
36	42.5	28.7	688	1 PYS2_PSEAE	Q06584 pseudomonas
37	42.5	28.7	776	1 VP4_ROTBU	P12474 bovine rota
38	42.5	28.7	776	1 VP4_ROTBU	O02945 equine rota
39	42.5	28.7	1752	1 RPB1_SCHPO	P36594 schizosacch
40	42	28.4	238	1 RISA_YEAST	P38145 saccharomyc
41	42	28.4	323	1 CDC2_EMENT	O00646 emeticella
42	42	28.4	339	1 NR12_ARATH	P32962 arabidopsis
43	42	28.4	346	1 NR1_ARATH	P32961 arabidopsis
44	42	28.4	450	1 CREO_ECOLI	P08369 escherichia
45	42	28.4	455	1 DEGO_ECOLI	P39099 escherichia

ALIGNMENTS

```

RESULT 1
PSB6_HUMAN STANDARD: PRT: 239 AA.
ID PSB6_HUMAN
AC P28072;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Proteasome subunit beta type 6 precursor (EC 3.4.25.1) (Proteasome
DE delta chain) (Macropain delta chain) (Multicatalytic endopeptidase
DE complex delta chain) (Proteasome subunit Y).
GN PSMB6 OR LMPY.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94345396; PubMed-8066462;
RA Akiyama K.-Y., Yokota K.-Y., Kagawa S., Shimbara N., Tamura T.,
RA Akioka H., Notthwang H.G., Noda C., Tanaka K., Ichihara A.;
RT "CDNA cloning and interferon gamma down-regulation of proteasomal
RT subunits X and Y.";
RT Science 265:1231-1234(1994).
RN [2]
RP SEQUENCE OF 60-239 FROM N.A., AND SEQUENCE OF 35-75; 80-110 AND
RP 210-233.
RX MEDLINE-91363412; PubMed-1888762;
RA Demartino G.N., Orth K., McCullough M.L., Lee L.W., Munn T.Z.,
RA Moomaw C.R., Dawson P.A., Slaughton C.A.;
RT "The primary structures of four subunits of the human, high-molecular-
RT weight proteinase, macropain (proteasome), are distinct but
RT homologous.";
RT Biochim. Biophys. Acta 1079:29-38(1991).
RN [3]
RP SEQUENCE OF 35-60.
RX MEDLINE-90167111; PubMed-2306472;
RA Lee L.W., Moomaw C.R., Orth K., McGuire M.J., Demartino G.N.,
RA Slaughton C.A.;
RT "Relationships among the subunits of the high molecular weight
RT proteinase, macropain (proteasome).";
RT Biochim. Biophys. Acta 1037:178-185(1990).
RN [4]
RP FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
RP WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
RP PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
RP SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
RP ACTIVITY. MAY CATALYZE BASAL PROCESSING OF INTRACELLULAR ANTIGENS.
RN [5]
RP CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
RP specificity.
RN [6]
RP PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
RP proteolytic pathway.
RN [7]
RP SUBUNIT: The proteasome is composed of at least 15 non identical
RP subunits which form a highly ordered ring-shaped structure. This
RP subunit can be displaced by the equivalent immune-specific subunit
RP PSMB9.
RN [8]
RP SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
RN [9]
RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B; ALSO KNOWN AS THE

```

```

CC      PROTEASOME B-TYPE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch)
CC      -----
DR      EMBL: D29012; BA006098.1; -
DR      EMBL: X61971; CAA43963.1; -
DR      PIR: S08188; S08188.
DR      PIR: S17522; S17522.
DR      MEROPS: T01.010; -
DR      SWISS-2DPAGE: P28072; HUMAN.
DR      PHCI-2DPAGE: P28072; -
DR      MIM: 600307; -
DR      InterPro: IPR001353; Proteasome.
DR      InterPro: IPR000243; Proteasome_B.
DR      Pfam: PF00227; Proteasome; 1.
DR      PRINTS: PR00141; PROTEASOME.
DR      PROSITE: PS00854; PROTEASOME_B; 1.
KW      Proteasome; Hydrolase; Protease; Zymogen.
FT      PROPEP 1 34
FT      CHAIN 35 239
FT      CONFLICT 145 145 G -> V (IN REF. 2).
SQ      SEQUENCE 239 AA; 25315 MW; 66EB9B96C685830D CRC64;

Query Match 36.1%; Score 53.5; DB 1; Length 239;
Best Local Similarity 43.5%; Pred. No. 1.7;
Matches 10; Conservative 6; Mismatches 2; Indels 5; Gaps 1;

QY 11 QGGGQFAPIG-----QMAING 28
Db 141 EGGGQYVPMGMMVRQSPAIIG 163

RESULT 2
YBDA_ECOLI
ID YBDA_ECOLI STANDARD; PRT; 416 AA.
AC P24077;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical membrane protein P43.
GN YBDA OR B0591.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92157868; PubMed=1838574;
RA Shea C.M., McIntosh M.A.;
RT "Nucleotide sequence and genetic organization of the ferric
RT enterobactin transport system: homology to other periplasmic binding
RT protein-dependent systems in Escherichia coli.";
RL Mol. Microbiol. 5:1415-1428(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92157867; PubMed=1787794;
RA Chenault S.S., Earhart C.F.;
RT "Organization of genes encoding membrane proteins of the Escherichia
RT coli ferriterobactin permease.";
RL Mol. Microbiol. 5:1405-1413(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;

```

```

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federapfel N., Hyman R., Kaiman S., Komp C., Kardi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch)
CC      -----
DR      EMBL: X57470; CAA40706.1; -
DR      EMBL: X59402; CAA42044.1; -
DR      EMBL: AE000164; AAC73692.1; -
DR      EMBL: U82598; ABA40790.1; -
DR      PIR: S16295; S16295.
DR      PIR: S16306; S16306.
DR      PIR: S14850; S14850.
DR      EcoGene: EG11104; Ybda.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 22 42
FT TRANSMEM 56 76
FT TRANSMEM 96 116
FT TRANSMEM 119 139
FT TRANSMEM 157 177
FT TRANSMEM 179 199
FT TRANSMEM 219 239
FT TRANSMEM 257 277
FT TRANSMEM 301 321
FT TRANSMEM 357 377
FT TRANSMEM 379 399
FT TRANSMEM 123 123
FT CONFLICT 123 123 F -> S (IN REF. 1).
SQ SEQUENCE 416 AA; 43282 MW; 64FCF9A9F9AC25A CRC64;

Query Match 33.4%; Score 49.5; DB 1; Length 416;
Best Local Similarity 46.2%; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 6; Indels 3; Gaps 1;

QY 2 AASDNFQLSGGGQGF--AIPIGAM 24
Db 248 ALADNMQMSAAQIGFLYAIPLGAAI 273

RESULT 3
YEGK_ECOLI
ID YEGK_ECOLI STANDARD; PRT; 253 AA.
AC P76395;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yegk.
GN YEGK OR B2072.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

```

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000297; AAC75133.1; -.
 DR Ecogene: BG14054; yegK.
 DR Hypothetical protein: Complete proteome.
 KM Hypothetical protein: Complete proteome.
 SO SEQUENCE 253 AA: 27208 MW: DC828DD5AD67E7E0 CRC64;

Query Match 33.1%; Score 49; DB 1; Length 253;
 Best Local Similarity 42.9%; Pred. No. 7.6;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 9 LSGGGGFAPIGAMWINGO 29
 DB 51 VSGGEGAMLVANMAVMSQ 71

RESULT 4
 ID BCHH_RHOCA STANDARD; PRT; 1194 AA.
 AC P26162;
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Magnesium-chelatase subunit H (Mg-Protochlorophyllin IX chelatase subunit
 DE H).
 GN BCHH.
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxID=1061;
 OX [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=SB1003;
 RX MEDLINE=93224465; PubMed=8385667;
 RA Burke D.H., Alberti M., Hearst J.E.;
 RA "bchNbh bacteriochlorophyll synthesis genes of Rhodobacter
 RT capsulatus and identification of the third subunit of
 RT light-independent protochlorophyllide reductase in bacteria and
 RT plants.";
 RT J. Bacteriol. 175:2414-2422(1993).
 RL J. Bacteriol. 175:2414-2422(1993).
 RN [2]
 RN SEQUENCE OF 1170-1194 FROM N.A.
 RP MEDLINE=90368552; PubMed=2203738;
 RX Yang Z., Bauer C.E.;
 RA "Rhodobacter capsulatus genes involved in early steps of the
 RT bacteriochlorophyll biosynthetic pathway.";
 RT J. Bacteriol. 172:5001-5010(1990).
 RL J. Bacteriol. 172:5001-5010(1990).
 CC -1- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
 CC A MAGNESIUM ION INTO PROTOCHLOROPHYRIN IX TO YIELD MG-
 CC PROTOCHLOROPHYRIN IX.
 CC -1- PATHWAY: LIGHT-INDEPENDENT BACTERIOCHLOROPHYLL BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNIT H FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

CC EMBL: 211165; AAA7524.1; -.
 DR EMBL: M34843; AAA26097.1; -.
 DR PIR: S17808; S17808.
 DR PIR: A36716; A36716.
 DR PIR: D49851; D49851.
 DR InterPro: IPR003672; COBN/Mg-chelatase.
 DR Pfam: PF02514; COBN-Mg_chel; 1.
 KM Photosynthesis: Bacteriochlorophyll biosynthesis.
 SO SEQUENCE 1194 AA: 129361 MW: 6341816A58774EE5 CRC64;

Query Match 33.1%; Score 49; DB 1; Length 1194;
 Best Local Similarity 52.0%; Pred. No. 35;
 Matches 13; Conservative 3; Mismatches 3; Indels 6; Gaps 2;

OY 4 SDNFOLSGGGFAPIGAMWING 28
 DB 848 SDNIR-SDGG-----PIGQALALMG 866

RESULT 5
 ID YCXA_EUGCR STANDARD; PRT; 350 AA.
 AC P31561;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 40.9 kDa protein in 16S rRNA 3' region (ORF350).
 DE Euglena gracilis.
 OS Euglenozoa; Euglenida; Euglenales; Euglena.
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
 OX NCBI_TaxID=3039;
 OX [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Z;
 RX MEDLINE=93347989; PubMed=8346031;
 RA Orsat B., Spielmann A., Stutz E.;
 RA Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A.,
 RA "Complete sequence of *Euglena gracilis* chloroplast DNA.";
 RT Nucleic Acids Res. 21:3537-3544(1993).
 RL [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Z;
 RC Schluegger B., Stutz E.;
 RA "The *Euglena gracilis* chloroplast genome: structural features of a
 RT DNA region possibly carrying the single origin of DNA replication.";
 RT Curr. Genet. 8:629-634(1984).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 211874; -; NOT_ANNOTATED_CDS.
 DR EMBL: X70810; CAA50140.1; -.
 DR PIR: S34557; S34557.
 DR PIR: S36933; S36933.
 KM Chloroplast: Hypothetical protein.
 SO SEQUENCE 350 AA: 40861 MW: 1AACAT7EC540FF51 CRC64;

Query Match 32.4%; Score 48; DB 1; Length 350;
 Best Local Similarity 56.2%; Pred. No. 14;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 15 GFAIPGAMATAGOI 30
 DB 315 GFVYPGQMSAPGOI 330

```

RESULT 6
ABC7_MOUSE
ID ABC7_MOUSE STANDARD: PRT; 694 AA.
AC 061102;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family B, member 7, mitochondrial (ATP-binding cassette transporter 7) (ABC transporter 7 protein)
DE (Fragment)
GN ABC7 OR ABC7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DRA/2;
RA MEDLINE=97288528; PubMed=9143506;
RA Savary S., Allikmets R., Denizot F., Luciano M.-F., Mattei M.-G., Dean M., Chimi G.;
RT Isolation and chromosomal mapping of a novel ATP-binding cassette transporter conserved in mouse and human.;
RL Genomics 41:275-278(1997).
CC -1- FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF HEME FROM THE MITOCHONDRIA TO THE CYTOSOL. PLAYS A CENTRAL ROLE IN THE MATURATION OF CYTOSOLIC IRON-SULFUR (FE/S) CLUSTER-CONTAINING PROTEINS (By similarity).
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial inner membrane (potential).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U43892; AAC53152.1; -.
CC MGD: MGI:109333; Abcd7.
CC DR InterPro: IPR003593; AAA.
CC DR InterPro: IPR001440; ABC_transporter_tmam.
CC DR InterPro: IPR003439; ABC_transporter.
CC DR InterPro: IPR001687; ATP_GTP_A.
CC DR Pfam: PF00664; ABC_membrane.1.
CC DR Pfam: PF00005; ABC_tran.1.
CC DR SMART: SM00382; AAA.1.
CC DR PROSITE: PS00211; ABC_TRANSPORTER.1.
CC DR ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane.
CC FT NON_TER 1 1
CC FT TRANSMEM 202 222 POTENTIAL.
CC FT TRANSMEM 233 253 POTENTIAL.
CC FT TRANSMEM 326 346 POTENTIAL.
CC FT NP_BIND 447 454 ATP (POTENTIAL).
CC FT SEQUENCE 694 AA; 76417 MW; A7AB89AE9AA9BID CRC64;

```

```

Query Match 32.1%; Score 47.5; DB 1; Length 694;
Best Local Similarity 37.1%; Pred. No. 33;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

```

```

QY 1 TASDN-FQLSGC-----GQFAIPICGMATAIG 28
DB 413 TVAFDNVHEFYEGQVYLVNGSVFEPAGKVAIVG 447

```

```

RESULT 7
HYPE_ALCEU
ID HYPE_ALCEU STANDARD: PRT; 351 AA.

```

```

AC P31905;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1993 (Rel. 38, Last annotation update)
DE Hydrogenase expression/formation protein hype.
GN HYPE.
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Plasmid megaplasmid PHG1.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
OC NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H16 / ATCC 17699;
RA MEDLINE=93356597; PubMed=8352644;
RA Dermede J., Eitlinger M., Friedrich B.;
RT Analysis of a plasmid gene region involved in formation of catalytically active hydrogenases in Alcaligenes eutrophus H16.;
RL Arch. Microbiol. 159:545-553(1993).
CC -1- SIMILARITY: BELONGS TO THE HYPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X70183; CAA49735.1; -.
CC DR PIR: S29979; S29979.
CC DR InterPro: IPR000728; AIRS-related.
CC DR Pfam: PF00586; AIRS.1.
CC DR Pfam: PF02769; AIRS_C.1.
CC KM Plasmid.
CC SQ SEQUENCE 351 AA; 36596 MW; 35D7291974B7BBBD CRC64;

```

```

Query Match 31.8%; Score 47; DB 1; Length 351;
Best Local Similarity 46.4%; Pred. No. 20;
Matches 13; Conservative 2; Mismatches 11; Indels 2; Gaps 1;

```

```

QY 2 AASDNFQLSGGQ--GFAIPICGMATA 27
DB 41 AARDNFWLRGNDQAAAFAMPAGARWVA 68

```

```

RESULT 8
T3RE_BACCE
ID T3RE_BACCE STANDARD: PRT; 412 AA.
AC P25241;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type III restriction-modification system BceI0987IP enzyme res (EC 3.1.21.5) (Fragment).
CC GN RES.
CC OS Bacillus cereus.
CC OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus.
CC OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10987;
RA MEDLINE=92267377; PubMed=1587478;
RA Hegde I.K., Karlstrom E.S., Lopez R., Kristensen T., Kolstoe A.-B.;
RT "A type-III DNA restriction and modification system in Bacillus cereus";
RL Gene 114:149-150(1992).
CC -1- FUNCTION: THIS PROTEIN CUT THE DNA OUTSIDE OF THE RECOGNITION SITE. MAY ALSO ACT AS A HELICASE INVOLVED IN UNWINDING DNA AT THE CLEAVAGE SITE. PROTEIN ONLY REQUIRED FOR RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME (BY SIMILARITY).
CC

```


RA Bishop D.F.;
 RT "Human ABC7 transporter: gene structure and mutation causing x-linked
 RT sideroblastic anemia with ataxia with disruption of cytosolic
 RT iron-sulfur protein maturation.";
 RL Blood 96:3256-3264(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 5-752 FROM N.A.
 RX MEDLINE=99098366; PubMed=9883897;
 RA Csere P., Lill R., Kispal G.;
 RT "Identification of a human mitochondrial ABC transporter, the
 RT functional orthologue of yeast Atmtp.";
 RL FEBS Lett. 441:266-270(1998).
 CC -! FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF HEME FROM THE
 CC MITOCHONDRION TO THE CYTOSOL. PLAYS A CENTRAL ROLE IN THE
 CC MATURATION OF CYTOSOLIC IRON-SULFUR (FE/S) CLUSTER-CONTAINING
 CC PROTEINS.
 CC -! SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
 CC -! SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (potential).
 CC -! DISEASE: DEFECTS IN ABC7 ARE THE CAUSE OF X-LINKED SIDEROBLASTIC
 CC ANEMIA WITH ATAXIA (ASAT). ASAT IS A RECESSIVE DISORDER
 CC CHARACTERIZED BY AN INFANTILE TO EARLY CHILDHOOD ONSET OF
 CC NONPROGRESSIVE CEREBELLAR ATAXIA AND MILD ANEMIA WITH HYPOCHROMIA
 CC AND MICROCYTOSIS.
 CC -! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUPERFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB005289; BA28861.1; -;
 DR EMBL: AF038950; AAC39865.1; -;
 DR EMBL: AF133659; AAD33045.1; -;
 DR EMBL: AF241887; AAK20173.1; -;
 DR EMBL: AF241872; AAK20173.1; JOINED.
 DR EMBL: AF241873; AAK20173.1; JOINED.
 DR EMBL: AF241874; AAK20173.1; JOINED.
 DR EMBL: AF241875; AAK20173.1; JOINED.
 DR EMBL: AF241876; AAK20173.1; JOINED.
 DR EMBL: AF241877; AAK20173.1; JOINED.
 DR EMBL: AF241878; AAK20173.1; JOINED.
 DR EMBL: AF241879; AAK20173.1; JOINED.
 DR EMBL: AF241880; AAK20173.1; JOINED.
 DR EMBL: AF241881; AAK20173.1; JOINED.
 DR EMBL: AF241882; AAK20173.1; JOINED.
 DR EMBL: AF241883; AAK20173.1; JOINED.
 DR EMBL: AF241884; AAK20173.1; JOINED.
 DR EMBL: AF241885; AAK20173.1; JOINED.
 DR EMBL: AF241886; AAK20173.1; JOINED.
 DR EMBL: BC006323; AAK06323.1; -;
 DR EMBL: AF078777; AAD47141.1; -;
 DR MIM: 300135; -;
 DR MIM: 301310; -;
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC_transporter_tmam.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00064; ABC_membrane; 1.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane;
 KW Transic peptide; Disease mutation.
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).

FT	CHAIN	?	752	ATP-BINDING CASSETTE, SUB-FAMILY B, MEMBER 7. POTENTIAL.
FT	TRANSMEM	260	280	POTENTIAL.
FT	TRANSMEM	291	311	POTENTIAL.
FT	TRANSMEM	383	403	POTENTIAL.
FT	NP_BIND	505	512	ATP (POTENTIAL).
FT	VARIANT	400	400	I -> M (IN ASAT).
FT	VARIANT	400	400	/FTID=VAR_009156.
FT	VARIANT	433	433	E -> K (IN ASAT; IMPAIRED MATURATION OF CYTOSOLIC FE/S PROTEINS).
FT	CONFLICT	56	56	/FTID=VAR_012640.
FT	CONFLICT	141	141	O -> OO (IN REF. 5).
FT	CONFLICT	258	258	A -> P (IN REF. 2).
FT	CONFLICT	271	276	R -> K (IN REF. 1).
FT	CONFLICT	281	282	LIPIN6 -> PLIPNHV (IN REF. 2).
FT	CONFLICT	290	290	VS -> LVS (IN REF. 2).
FT	CONFLICT	293	297	G -> C (IN REF. 2).
FT	CONFLICT	315	315	FALVT -> LLGN (IN REF. 2).
FT	CONFLICT	320	324	R -> G (IN REF. 1).
FT	CONFLICT	346	346	TEMNK -> LEIDQ (IN REF. 2).
FT	CONFLICT	542	542	F -> I (IN REF. 1).
FT	CONFLICT	542	542	E -> V (IN REF. 6).
SO	SEQUENCE	752 AA;	82641 MW;	BLFFA57ABD24FB90 CRC64;

Query Match 31.4%; Score 46.5; DB 1; Length 752;
 Best Local Similarity 37.1%; Pred No. 50;
 Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

QY 1 TAASDN--FOLSOG-----GQGFAPIGQMAIAG 28
 Db 471 TVAFDNVHFYEIEGQKVLGISFEVPARKVAIVG 505

RESULT 11
 PMP9.CHLPN
 ID PMP9.CHLPN STANDARD; PRT; 928 AA.
 AC Q92398;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable outer membrane protein pmp9 precursor (Polymorphic membrane protein 9) (outer membrane protein 10).
 GN PMP9 OR OMP10 OR CPN0447 OR CP0306.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VR1310;
 RX MEDLINE=20007584; PubMed=10539856;
 RA Christiansen G., Boesen T., Hjerno K., Daugaard L., Mygind P.,
 RA Madsen A.S., Knudsen K., Falk E., Birkedal S.;
 RT "Molecular biology of Chlamydia pneumoniae surface proteins and their
 RT role in immunopathogenicity.";
 RL Am. Heart J. 138:S491-S495(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger U., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uitterlind T., Berry K., Baas S.,
 RA Linher M., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;


```

RT      *Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT      pneumoniae AR39."
RL      Nucleic Acids Res. 28:1397-1406(2000).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN-J138;
RA      MEDLINE=20330349; PubMed=10871362;
RA      Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA      Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT      "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT      from Japan and CWL029 from USA."
RL      Nucleic Acids Res. 28:2311-2314(2000).
CC      -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC      (POTENTIAL).
CC      -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
-----
CC      EMBL: AJ13034; CAB37069.1; -
DR      EMBL: AE001628; AAD18591.1; -
DR      EMBL: AE002192; AAF8163.1; -
DR      EMBL: AP002546; BAA98655.1; -
DR      TIGR: CP0306; -
DR      InterPro: IPR003368; DUF145.
DR      InterPro: IPR003357; OMP.
DR      Pfam: PF02415; DUF145; 1.
DR      Pfam: PF02385; OMP; 1.
DR      Outer membrane; Signal; Multigene family; Complete proteome.
FT      SIGNAL 1 26
FT      CHAIN 27 928 PROBABLE OUTER MEMBRANE PROTEIN PMP9.
FT      SEQUENCE 928 AA; 98332 MW; 58910A8F04F12219 CRC64;
SQ
Query Match 31.4%; Score 46.5; DB 1; Length 928;
Best Local Similarity 39.4%; Pred. No. 62;
Matches 13; Conservative 3; Mismatches 8; Indels 9; Gaps 1;
OY      4 SDFOLSGGGGF-----AIPGQMAIA 27
DB      293 TDNLVSSGGPTLFKNSAIDTAAPLGAIA 325
-----
RESULT 12
CTPI_MYCLE STANDARD; PRT: 1609 AA.
AC      053114;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Probable cation-transporting ATPase I (EC 3.6.3.-).
GN      CTPI OR MD2671 OR MCB1913.02.
OS      Mycobacterium leprae.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1769;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-TN;
RA      MEDLINE=21128732; PubMed=11234002;
RA      Cole S.T., Eijmeier K., Parkhill J., James K.D., Thomson N.R.,
RA      Wheeler P.R., Honore N., Garnier T., Churches C., Harris D.,
RA      Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R.M., Devlin K., Dutney S., Feltwell T., Fraser A., Hamlin N.,
RA      Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA      Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA      Rutter S., Seeger K., Simon S., Simmonds K., Skelton J., Squares R.,
RA      Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

```

```

RA      Barrell B.G.;
RT      "Massive gene decay in the leprosy bacillus."
RL      Nature 409:1007-1011(2001).
CC      -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC      (E1-E2 ATPASES).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
-----
CC      EMBL: AL022118; CAI17934.1; -
DR      EMBL: AL583926; CAC32203.1; -
DR      Leptoma; M12671; -
DR      InterPro: IPR001757; E1-E2_ATPase.
DR      InterPro: IPR001454; Hydrolase.
DR      Pfam: PF00122; E1-E2_ATPase; 1.
DR      Pfam: PF00702; Hydrolase; 1.
DR      PRINTS: PR00119; CATATPASE.
DR      PROSITE: PS00154; ATPASE_E1_E2; 1.
KW      Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW      Complete proteome.
FT      TRANSMEM 30 50
FT      TRANSMEM 176 196
FT      TRANSMEM 238 258
FT      TRANSMEM 357 377
FT      TRANSMEM 641 661
FT      TRANSMEM 673 693
FT      TRANSMEM 778 798
FT      TRANSMEM 921 941
FT      TRANSMEM 969 989
FT      TRANSMEM 997 1017
FT      TRANSMEM 1396 1416
FT      TRANSMEM 1426 1446
FT      TRANSMEM 1542 1562
FT      TRANSMEM 1573 1593
FT      MOD_RES 1053 1053
FT      METAL 1335 1335
FT      METAL 1339 1339
FT      SEQUENCE 1609 AA; 166871 MW; 58FA2079905E3995 CRC64;
SQ
Query Match 31.1%; Score 46; DB 1; Length 1609;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
OY      12 GGGGFAIPGQMAIAGOI 30
DB      388 GQQAFAATLGRGLANAGOL 406
-----
RESULT 13
CARA_MYCTU STANDARD; PRT: 376 AA.
ID      CARA_MYCTU
AC      P71811;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Carbamoyl-phosphate synthase glutamine small chain (EC 6.3.5.5) (Carbamoyl-
DE      phosphate synthetase glutamine small chain).
GN      CARA OR RV1383 OR MT1427 OR MTCY02B12.17.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-H37RV;

```

RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Deakin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jorgensen K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
 RA Sulston J.B., Taylor K., Whitehead S., Barrall B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Minkula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains.";
 Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + glutamate + carbamoyl phosphate.
 CC -I- PATHWAY: INVOLVED IN BOTH ARGININE AND PYRIMIDINE BIOSYNTHESIS
 CC -I- SUBUNIT: COMPOSED OF TWO CHAINS: THE SMALL (OR GLUTAMINE) CHAIN
 CC PROMOTES THE HYDROLYSIS OF GLUTAMINE TO AMMONIA, WHICH IS USED
 CC BY THE LARGE (OR AMMONIA) CHAIN TO SYNTHESIZE CARBAMOYL PHOSPHATE.
 CC -I- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/
 CC or_send_an_email_to_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).
 CC -----
 DR EMBL; Z81011; CAB02644.1; -;
 DR EMBL; AE007014; AAK45692.1; ALT_INIT.
 DR HSSP; P00907; ICE8.
 DR TIGR; MT1427; -;
 DR TIGR; MT1427; RV1383; -;
 DR InterPro; IPR001317; CPS_GATase.
 DR InterPro; IPR002474; CPSase_sm_chain.
 DR InterPro; IPR000991; GATase_1.
 DR Pfam; PF00988; CPSase_sm_chain; 1.
 DR Pfam; PF00117; GATase; 1.
 DR PRINTS; PR00099; CPSGATASE.
 DR PRINTS; PR00096; GATASE.
 DR PROSITE; PS00442; GATASE_TYPE_1; 1.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase;
 KW Complete proteome.
 FT DOMAIN 1 180 CPSASE.
 FT ACT_SITE 181 376 GLUTAMINE AMIDOTRANSFERASE.
 FT ACT_SITE 260 260 GATASE (BY SIMILARITY).
 FT SEQUENCE 376 AA; 39766 MW; Df057465F47AE46D CRC64.

Query Match	30.7%	Score 45.5;	DB 1;	Length 376;
Best Local Similarity	44.0%	Pred. No. 35;		
Matches 11; Conservative	3;	Mismatches 10;	Indels 1;	Gaps 1

```

OY      1  TAASDNFOL-SQGGQGFALPIQAM  24
        || : || : || || || :
Db      299 TAQNHGFALQGEAGQSFATPEGPAV  323

```

RESULT ' 14	
REBX_SHIDY	
ID REBX_SHIDY	STANDARD;
	PRT; 396 AA

AC	Q03563;
AD	01-FEB-1995 (rel. 31, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Putative O-antigen transporter.
GN	PEPX.
OS	Shigella dysenteriae.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Shigella.
OX	NCBI_TaxID=622;
RP	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94018634; PubMed=7692219;
RA	Klena J.D., Schmidtman C.A.;
RT	"Function of the rfb gene cluster and the rfe gene in the synthesis
RT	of O antigen by Shigella dysenteriae 1.";
RL	Mol. Microbiol. 9:393-402(1993).
CC	-1- FUNCTION: MAY BE INVOLVED IN THE TRANSLLOCATION PROCESS OF THE
CC	NASENT O-POLYSACCHARIDE MOLECULES AND/OR ITS LIGATION TO LIPID
CC	A CORE UNITS.
CC	-1- PATHWAY: O ANTIGEN SYNTHESIS IN LIPIDPOLYSACCHARIDE BIOSYNTHESIS.
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC	(potential).
CC	-1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE SYNTHASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its use
CC	by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by, and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@sib-sib.ch).
CC	-----
DR	EMBL, L07293; AAA16934.1; -.
DR	PIR, S34963; S34963.
DR	InterPro: IPR002797; Polysacc_synt.
DR	Pfam; PF01943; Polysacc_synt; 1.
KW	Lipopolysaccharide biosynthesis; Transport; Transmembrane;
KW	Inner membrane.
FT	TRANSMEM 5 25 POTENTIAL.
FT	TRANSMEM 32 52 POTENTIAL.
FT	TRANSMEM 82 102 POTENTIAL.
FT	TRANSMEM 107 127 POTENTIAL.
FT	TRANSMEM 137 157 POTENTIAL.
FT	TRANSMEM 163 183 POTENTIAL.
FT	TRANSMEM 211 231 POTENTIAL.
FT	TRANSMEM 286 306 POTENTIAL.
FT	TRANSMEM 322 342 POTENTIAL.
FT	TRANSMEM 348 368 POTENTIAL.
FT	TRANSMEM 372 392 POTENTIAL.
SO	SEQUENCE 396 AA; 44309 MW; A1C5E06865CDFAEE CRC64;

	Query	March	Similarity	Score	DB	Length	396;
	Best	Local	5	52.98;	Pred.	No.	43;
	Matches	9;	Conservative	2;	Mismatches	6;	Indels
	0;	Gaps	0;				
QY	8	QLSGGGGAFAIPICGAM	24				
		..					
db	246	KLRMAAGLIRPICGAV	262				

```
QY      8 QLSQGQGFAPIGAM 24
          :| | | | | | | :
Db     246 KLRMAAQGLPIGOAV 262
```

RESULT	15
60IM_HAEIN	
ID	60IM_HAEIN
AC	P44973
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	60 kDa inner-membrane protein homolog.
GN	HI1001
OS	Haemophilus influenzae.
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellales; Pasteurellaceae;

```

OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-RD / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Giodè A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.":
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE OXAL / 60 KDA IMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32781; AAC22663.1; -.
DR TIGR: H11001; -.
DR InterPro: IPR001708; 60KD_inner_MP.
DR Pfam: PF02096; 60KD_IMP_1.
DR PRINTS: PR00701; 60KDINNERMP.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 495 515 POTENTIAL.
SQ SEQUENCE 541 AA; 61239 MW; 9406398B334FAEDE CRC64;

```

```

Query Match 30.4%; Score 45; DB 1; Length 541;
Best Local Similarity 37.5%; Pred. No. 59;
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 5 DNFOLSGGGGFAIP1 20
DB 139 DNFKLAEGGESLSVPL 154

```

Search completed: August 13, 2002, 10:42:45
Job time: 388 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:47 ; Search time 43.59 seconds
(without alignments)
119.061 Million cell updates/sec

Title: US-09-684-215A-17
Perfect score: 148
Sequence: 1 TAASDNFQLSQGCGFAIPGMAIAIGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	355	16 007175	007175 mycobacteri
2	104.5	70.6	361	2 050320	050320 mycobacteri
3	98.5	66.6	354	16 09CCY9	09CCY9 mycobacteri
57	38.5	780	4 096C66		096C66 homo sapien
5	37.8	296	16 09A982		09A982 caulobacter
6	37.2	464	16 053886		053886 mycobacteri
7	37.2	542	2 09FBR3		09FBR3 streptomyc
8	36.5	125	4 09BV14		09BV14 homo sapien
9	36.5	382	16 09CD67		09CD67 mycobacteri
10	36.5	407	4 096KSO		096KSO homo sapien
11	36.5	452	2 092556		092556 mycobacteri
12	36.5	452	16 P73354		P73354 synecocyst
13	36.5	419	11 099MT0		099MT0 mus musculu
14	35.8	419	11 091YE2		091YE2 mus musculu
15	35.1	519	2 093J30		093J30 streptomyc
16	35.1	630	11 099JIV6		099JIV6 mus musculu

17	51.5	34.8	820	10 095XB5	095XB5 arabidopsis
18	51	34.5	369	16 092839	092839 chlamydia p
19	50.5	34.1	1000	16 09PCD0	09PCD0 xylella fas
20	50	33.8	362	2 09S2R5	09S2R5 streptomyc
21	50	33.8	487	10 09LPN6	09LPN6 arabidopsis
22	49.5	33.4	358	10 096520	096520 arabidopsis
23	49.5	33.4	358	10 043734	043734 arabidopsis
24	49.5	33.4	333	16 09KJ0	09KJ0 bacillus ha
25	49.5	33.4	571	2 09L4P0	09L4P0 staphylococ
26	49.5	33.4	662	16 099UD4	099UD4 staphylococ
27	49.5	33.4	784	12 091B33	091B33 equine rhin
28	49	33.1	275	10 091IX7	091IX7 oryza sativ
29	49	33.1	340	10 023532	023532 arabidopsis
30	49	33.1	394	16 P72780	095SP3 arabidopsis
31	49	33.1	591	16 0913V3	0913V3 pseudomonas
32	49	33.1	665	16 092E08	092E08 listeria in
33	49	33.1	966	5 095XN3	095XN3 caenorhabdi
34	49	33.1	2553	5 022860	022860 listeria in
35	49	32.4	348	16 0926M0	0926M0 escherichia
36	48	32.4	374	2 09S519	09S519 salmonella
37	48	32.4	374	2 093Q24	093Q24 rhizobium 1
38	48	32.4	407	16 098J66	098J66 azotobacter
39	48	32.4	473	2 044476	044476 pseudomonas
40	48	32.4	481	2 09AOD1	09AOD1 lactococcus
41	48	32.4	506	16 09CFT4	069854 streptomyc
42	48	32.4	697	2 069854	09RX85 delnoccocus
43	48	32.4	822	16 09RX95	09RX95 delnoccocus
44	48	32.4	341	10 093104	093104 arabidopsis
45	47.5	32.1			

ALIGNMENTS

RESULT	ID	007175	PRELIMINARY:	PRT:	355 AA.
AC	007175:				
DT	01-JUL-1997 (TREMBLrel. 04, Created)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	HYPOTHETICAL 34.9 KDA PROTEIN.				
GN	PEPA OR RV0125 OR MTC1418B.07.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornshy T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
CC	-I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.				
DR	EMBL; 296071; CAB09453.1; -				
DR	MEROFS; S01.UFC;				
DR	TuberculinSt; RV0125; -				
DR	InterPro; IPR001478; PDZ.				
DR	InterPro; IPR001254; Trypsin.				
DR	Pfam; PF00595; PDZ. 1.				
DR	Pfam; PF00089; trypsin. 1.				
DR	SMART; SM00226; PDZ. 1.				
DR	PROSITE; PS50106; PDZ. 1.				
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.				

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN.L.
KW Complete proteome: Hydrolase; Hypothetical protein; Serine protease.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match
Best Local Similarity 100.0%; Score 148; DB 16; Length 355;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASNFOLSGGGGFAIPIGQAMAIAGQI 30
DB 224 TAASNFOLSGGGGFAIPIGQAMAIAGQI 253

RESULT 2
ID 050320 PRELIMINARY; PRT; 361 AA.
AC 050320;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 34KDA PROTEIN PRECURSOR
OS Mycobacterium paratuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JD88/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterization of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
DR EMBL: Z23092; CAAB0638.1;
DR MEROPS: S01.UPC;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 36
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match
Best Local Similarity 70.6%; Score 104.5; DB 2; Length 361;
Matches 21; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

OY 1 TAASNFOLSGGGGFAIPIGQAMAIAGQI 30
DB 231 TAATDSYKMS-GGOGFAIPIGRAMAVANQI 259

RESULT 3
ID 09CCY9 PRELIMINARY; PRT; 354 AA.
AC 09CCY9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROBABLE SECRETED SERINE PROTEASE.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;

RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean A., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1 SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AL583926; CAC32191.1;
DR MEROPS: S01.UPC;
DR Leproma; ML2659;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR000126; Ser.proteas_V8.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9E4A4 CRC64;

Query Match
Best Local Similarity 66.7%; Score 98.5; DB 16; Length 354;
Matches 20; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 1 TAASNFOLSGGGGFAIPIGQAMAIAGQI 30
DB 224 TAATDNYKML-IGGGFAIPIGQAMEYVGA 252

RESULT 4
ID 096C66 PRELIMINARY; PRT; 780 AA.
AC 096C66;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO RIKEN CDNA 1810073N04 GENE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC014642; AAH14642.1;
SQ SEQUENCE 780 AA; 83324 MW; E529A228C842312 CRC64;

Query Match
Best Local Similarity 38.5%; Score 57; DB 4; Length 780;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 3 ASDNFOLSGGGGFAIPITGQA 23
DB 474 APEAQLDRPGGIAVPVGEA 494

RESULT 5
ID 09A9B2 PRELIMINARY; PRT; 296 AA.

```
AC 09A982:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOThETICAL PROTEIN CC1082.
GN CC1082.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OC NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernm W.C., Feldhlym T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson R.J., Durkin A.S., Gwinn M.L., Hatt D.H.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Hatt D.H.,
RA Kolony J.F., Smt J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uteback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005785; AAK2306.1; -.
DR TIGR; CC1082; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 296 AA; 31983 MW; 2EB4EBA376823924 CRC64;

Query Match 37.8%; Score 56; DB 16; Length 296;
Best Local Similarity 37.5%; Pred. No. 5.1;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 4 SDNFOLSGGGGFAIPGQAMAIYA 27
| : : | : : | : : | : : |
Db 159 SVEFIMTGSVGMVVPVGAIAVA 182

RESULT 6
053896 PRELIMINARY; PRT; 464 AA.
AC 053896:
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE SERINE PROTEASE.
GN RV0983 OR MTW04.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares K.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -; SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL021999; CAAL1582.1; -.
DR MEROPS; S01.0PC; -.
DR Tuberculist; RV0983; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR HydroLase; Protease; Serine protease.
KW SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;

Query Match 37.2%; Score 55; DB 2; Length 542;
Best Local Similarity 44.4%; Pred. No. 14;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 4 SDNFOLSGGGGFAIPGQAMAIYGOI 30
| : : | : : | : : | : : |
Db 412 SDDGAGSIGIGFAIPINOGKRAVEEL 438
```

```
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; HydroLase; Protease; Serine protease.
SQ SEQUENCE 464 AA; 46452 MW; AE93BRC53E1BC8F CRC64;

Query Match 37.2%; Score 55; DB 16; Length 464;
Best Local Similarity 40.6%; Pred. No. 12;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TASNDFOLSGG--GGFAIPGQAMAIYGOI 30
| : : | : : | : : | : : |
Db 337 TLGADSDAAGSGSIGIGFAIPVDQAKRIADEL 368

RESULT 7
09FBK9 PRELIMINARY; PRT; 542 AA.
AC 09FBK9:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE PROTEASE.
GN SCP8.12.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 mb streptomycetes coelicolor A3(2) chromosome.";
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL390975; CAC01350.1; -.
DR MEROPS; S01.0PC; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR01217; FRICHEXTENS.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW HydroLase; Protease; Serine protease.
SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;
```


DR MEROPS; S01.UPC; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50240; TRYPsin.DOM; 1.
 DR HydroLase: Protease; Serine protease.
 KW SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 36.5%; Score 54; DB 2; Length 452;
 Best Local Similarity 44.8%; Pred. No. 16;
 Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 2 AASDNFQLSGGGGFAIPGAMAIAGOI 30
 DB 328 ADSGDAQSGSIGLGFALPVDQAKRIADEL 356

RESULT 12
 ID P73354 PRELIMINARY; PRT; 452 AA.
 AC P73354;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE SERINE PROTEASE HTRA.
 GN HTRA OR SLR1204.
 OS Synecchocystis sp. (strain PCC 6803).
 OC Synecchocystis: Cyanobacteria; Chroococcales; Synecchocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
 DR EMBL: D90905; BAA17385.1; -.
 DR MEROPS; S01.273; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; Trypsin; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50240; TRYPsin.DOM; 1.
 DR Complete proteome: HydroLase; Protease; Serine protease.
 KW SEQUENCE 452 AA; 47656 MW; AED4AFB1AC472361 CRC64;

Query Match 36.5%; Score 54; DB 16; Length 452;
 Best Local Similarity 50.0%; Pred. No. 16;
 Matches 15; Conservative 2; Mismatches 9; Indels 4; Gaps 1;

QY 1 TAASDNFQLSGGGGFAIPGAMAIAGOI 30
 DB 312 TAIIONAQ-----GIGFAIPINRAQETIAQOI 337

RESULT 13
 ID Q99MI0 PRELIMINARY; PRT; 419 AA.
 AC Q99MI0;

DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CELL GROWTH REGULATOR FALKOR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C.
 RA Erez N., Gudkov A.V., Peles E., Cohen I.R., Rotter V.;
 RT "Falkor is a novel cell growth regulator, isolated from a MEF-derived
 RT GSE library."
 RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF340231; AAK37525.1; -.
 KW SEQUENCE 419 AA; 45108 MW; 56891872CAFBAB2B9 CRC64;

Query Match 35.8%; Score 53; DB 11; Length 419;
 Best Local Similarity 42.9%; Pred. No. 20;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 AASDNFQLSGGGGFAIPGQ 22
 DB 394 AARDKYQLASGQKGVQVPSQ 414

RESULT 14
 ID Q91YE2 PRELIMINARY; PRT; 419 AA.
 AC Q91YE2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE EGLN2 PROTEIN.
 GN EGLN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21458562; PubMed=11574160;
 RA Taylor M.S.;
 RT "Characterisation and comparative analysis of the EGLN gene family."
 RL Gene 275:125-132(2001).
 DR EMBL: AJ310547; GAC42516.1; -.
 KW SEQUENCE 419 AA; 44937 MW; ED9A600BBB5886BF CRC64;

Query Match 35.8%; Score 53; DB 11; Length 419;
 Best Local Similarity 42.9%; Pred. No. 20;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 AASDNFQLSGGGGFAIPGQ 22
 DB 394 AARDKYQLASGQKGVQVPSQ 414

RESULT 15
 ID Q93J30 PRELIMINARY; PRT; 519 AA.
 AC Q93J30;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE PROTEASE.
 GN SCBAC25E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;

AC	AAAM01143;
XX	
DT	04-OCT-2001 (first entry)
XX	
DE	Mycobacterium tuberculosis antigen Ra12 first 30 amino acids.
XX	
KM	Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW	cytostatic; gene therapy; metastasis.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	WO200151633-A2.
XX	
PD	19-JUL-2001.
XX	
PF	16-JAN-2001; 2001WO-US01574.
XX	
PR	14-JAN-2000; 2000US-0483672.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI	Kalos MD, Fanger GR, Day CH, Retter MM, Stolk JA, Skeiky YAW,
PI	Wang A, Neagher MJ;
XX	
DR	WPI; 2001-425873/45.
XX	
PT	New polynucleotide encoding a prostate-specific protein, for
PT	diagnosing, monitoring and treating prostate cancer in a patient and
PT	for use in vaccines -
XX	
PS	Example 17; Page 394; 543pp; English.
XX	
CC	The present invention describes polynucleotide sequences (I) which encode
CC	prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC	and can be used in vaccine production and gene therapy. (I), (II),
CC	antibodies to (II), fusion proteins comprising (II), and isolated
CC	T cells prepared using (I) or (II) are used treat cancer in a patient.
CC	(I) and the antibodies are also used in the detection of cancer in a
CC	patient. The cancer that is diagnosed or treated is particularly
CC	prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC	(I) can be used for monitoring the progression of cancer in a patient.
CC	(I) and (II) can also be used to improve diagnostic and therapeutic
CC	methods for prostate cancer. They can indicate the level of metastasis
CC	as well as the prostate volume. AAH93357 to AAH93944 and AAU01115 to
CC	AAAU01318 represent polynucleotide and amino acid sequences used in the
CC	exemplification of the present invention.
SQ	
	Sequence 30 AA:
	Query Match 100.0%; Score 148; DB 22; Length 30;
	Best Local Similarity 100.0%; Pred. No. 2,7e-15;
	Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	
	1 TAASDNFOLSGGOGFAPIPGAMVIAIGI 30
Dd	1 taasdntfslsggggafalpgqamalaqt 30
RESULT 3	
AAG99028	
ID	AAG99028 standard; Protein; 30 AA.
XX	
AC	AAG99028;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Mycobacterium tuberculosis antigen Ra12 1st 30 amino acid sequence.
KM	Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW	chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW	prostate specific antigen; PSA.

XX	Mycobacterium tuberculosis.
OS	
PX	WO200134802-A2.
PN	
PD	17-MAY-2001.
PP	
PF	09-NOV-2000; 2000MO-US30904.
XX	
PR	12-NOV-1999; 99US-0439313.
XX	
PR	18-NOV-1999; 99US-0443686.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PX	Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SC;
PI	Kalos MD, Relter MW, Stolk JA, Day CH, Skelky YAW, Wang A;
XX	
DR	WPI; 2001-308785/32.
XX	
PT	Isolated polypeptide comprising at least an immunogenic portion of a
PT	prostate-specific protein, useful in the diagnosis and therapy of
PT	prostate cancer -
PS	
XX	
PS	Example 17; Page 294; 325pp; English.
XX	
CC	The present invention describes an isolated polypeptide (P1) comprising
CC	at least an immunogenic portion of a prostate-specific protein, or its
CC	variant. Also described are polynucleotides (N1) encoding (P1), (P1) and
CC	(N1) have cytostatic activity and can be used in vaccine production.
CC	The polypeptides, nucleic acids and antibodies from the present
CC	invention are useful in the diagnosis and therapy of prostate cancer.
CC	Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC	in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC	region. Prostate specific antigen (PSA) P501S was located on
CC	chromosome 1. AAH84671 to AAH85143 and AAC99000 to AAC99077 represent
CC	polynucleotide and polypeptide sequences used in the exemplification
CC	of the present invention.
XX	
SO	Sequence 30 AA:
QY	
Query Match	100.0%; Score 148; DB 22; Length 30;
Best Local Similarity	100.0%; Pred. No. 2,7e-15;
Matches 30; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 TAAADNFQLSGGGGGAIRPGAMAIAGCI 30
Db	1 taasdnfqlsgggggafrapdigamaiaagci 30

Search completed: August 13, 2002, 10:59:49
Job time: 77 sec

OY	1	TAASDNFQLSGGGFAIPGQMAAIRGQI	30
Db	1	taasdnfqlsgsggfaiprgamalaqql	30
RESULT 3			
AAG99028	ID	AAG99028 standard; Protein; 30 AA.	
XX	AC		
XX	AC	AAG99028;	
XX	DT	25-SEP-2001 (first entry)	
XX	DE	Mycobacterium tuberculosis antigen Ra12 1st 30 amino acid sequence.	
XX	KM	Human: prostate cancer; therapy; diagnosis; cat eye syndrome; chromosome 22q11.2; prostate-specific protein; chromosome 1;	
KM	KW	prostate specific antigen; PSA	
KW			

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:32 ; Search time 14.11 Seconds
(without alignments)
204.300 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148

Sequence: 1 TAASDNFQLSQGGGFAPIIGQAMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 30
Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_71:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match length	ID	Description

No matches found

Search completed: August 13, 2002, 10:59:14
Job time: 42 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:59:17 ; Search time 10.19 Seconds
(without alignments)
113.993 Million cell updates/sec

Title: US-09-684-215a-17

Perfect score: 148
Sequence: 1 TAASDNFQLSQGGGFAIPICGAMAIAQOI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 30
Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
---------------	-------	----------------	--------	----	-------------

No matches found

Search completed: August 13, 2002, 11:02:53
Job time: 216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:57 ; Search time 23.81 seconds
(without alignments)
217.969 Million cell updates/sec

Title: US-09-684-215A-17
Perfect score: 148
Sequence: 1 TAASDNFQLSGGCGFAIPICQAMAIACQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 30
Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 100 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_minc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: August 13, 2002, 11:02:36
Job time: 219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:11 ; Search time 20.7 Seconds
(without alignments)
35.399 Million cell updates/sec

Title: US-09-684-215A-17
Perfect score: 148
Sequence: 1 TAASDNFOLSGGCGFAIPICGAMAIAGOI 30

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	30	4 US-09-439-313-484	Sequence 484, Appl
2	148	100.0	132	4 US-08-818-112-66	Sequence 66, Appl
3	148	100.0	132	4 US-08-818-111-67	Sequence 67, Appl
4	148	100.0	132	4 US-09-056-556-66	Sequence 66, Appl
5	148	100.0	355	4 US-08-818-112-79	Sequence 79, Appl
6	148	100.0	355	4 US-08-818-111-80	Sequence 80, Appl
7	148	100.0	355	4 US-09-056-556-79	Sequence 79, Appl
8	55	37.2	580	4 US-08-818-112-75	Sequence 75, Appl
9	55	37.2	580	4 US-08-818-111-76	Sequence 76, Appl
10	55	37.2	580	4 US-09-056-556-75	Sequence 75, Appl
11	47.5	32.1	694	2 US-08-895-522-3	Sequence 3, Appl
12	47.5	32.1	694	3 US-09-195-391-3	Sequence 3, Appl
13	46.5	31.4	747	2 US-08-895-522-1	Sequence 1, Appl
14	46.5	31.4	747	3 US-09-195-391-1	Sequence 1, Appl
15	46	31.1	344	2 US-08-846-762-90	Sequence 90, Appl
16	44	29.7	460	4 US-09-199-637A-132	Sequence 132, Appl
17	43.5	29.4	776	1 US-07-603-133B-18	Sequence 18, Appl
18	43	29.1	1528	1 US-08-463-092B-6	Sequence 6, Appl
19	43	29.1	1528	2 US-08-463-092B-6	Sequence 6, Appl
20	43	29.1	1528	2 US-08-460-907B-6	Sequence 6, Appl
21	43	29.1	1528	3 US-08-463-179A-6	Sequence 6, Appl
22	43	29.1	1528	3 US-08-461-384B-6	Sequence 6, Appl
23	42.5	28.7	248	4 US-09-091-219-3	Sequence 3, Appl
24	42.5	28.7	2232	4 US-09-091-219-25	Sequence 25, Appl
25	42.5	28.7	2247	2 US-09-091-219-2	Sequence 2, Appl
26	42	28.4	323	2 US-08-874-347-21	Sequence 21, Appl
27	42	28.4	323	3 US-09-093-522-21	Sequence 21, Appl

28	42	28.4	547	1 US-08-340-203A-3	Sequence 3, Appl
29	42	28.4	547	2 US-08-452-567-3	Sequence 3, Appl
30	42	28.4	547	2 US-08-452-427-3	Sequence 3, Appl
31	42	28.4	547	3 US-09-085-407-3	Sequence 3, Appl
32	42	28.4	1243	2 US-08-557-139-2	Sequence 2, Appl
33	42	28.4	1334	2 US-08-996-545-2	Sequence 2, Appl
34	42	28.4	1334	4 US-09-328-320-2	Sequence 2, Appl
35	42	28.4	1531	1 US-08-141-893-2	Sequence 2, Appl
36	42	28.4	1531	1 US-08-463-092B-2	Sequence 2, Appl
37	42	28.4	1531	1 US-08-463-092B-2	Sequence 2, Appl
38	42	28.4	1531	2 US-08-462-109A-2	Sequence 2, Appl
39	42	28.4	1531	2 US-08-462-109A-4	Sequence 2, Appl
40	42	28.4	1531	2 US-08-460-907B-2	Sequence 2, Appl
41	42	28.4	1531	2 US-08-460-907B-4	Sequence 2, Appl
42	42	28.4	1531	3 US-08-463-179A-2	Sequence 2, Appl
43	42	28.4	1531	3 US-08-463-179A-4	Sequence 2, Appl
44	42	28.4	1531	3 US-08-461-384B-2	Sequence 2, Appl
45	42	28.4	1531	3 US-08-461-384B-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-09-439-313-484
; Sequence 484, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqun
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439.313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-439-313-484

Query Match          100.0%; Score 148; DB 4; Length 30;
Best local Similarity 100.0%; Pred. No. 1e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TAASDNFOLSGGCGFAIPICGAMAIAGOI 30
Db      1 TAASDNFOLSGGCGFAIPICGAMAIAGOI 30

RESULT 2
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
```

APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-66

Query Match 100.0%; Score 148; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 6e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGFAIPIGQAMAIAGOI 30
Db 1 TAASDNFOLSGGGFAIPIGQAMAIAGOI 30

RESULT 3
US-08-818-111-67
Sequence 67, Application US/08818111
Patent No. 633852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111

FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-67

Query Match 100.0%; Score 148; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 6e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGFAIPIGQAMAIAGOI 30
Db 1 TAASDNFOLSGGGFAIPIGQAMAIAGOI 30

RESULT 4
US-09-056-556-66
Sequence 66, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
TUBERCULOSIS
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-66

Query Match 100.0%; Score 148; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 6e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TRE


```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match          100.0%; Score 148; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNFOLSGG--GQGFAPIGAMAIAGQI 30
Db 224 TAAADNFOLSGG--GQGFAPIGAMAIAGQI 253

RESULT 8
US-08-818-112-75
; Sequence 75, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; TITLE OF INVENTION: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-75

Query Match          37.2%; Score 55; DB 4; Length 580;
Best Local Similarity 40.6%; Pred. No. 2.5;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TAAADNFOLSGG--GQGFAPIGAMAIAGQI 30
```

```

Db 453 TLGADSDAQSGSIGLGFAPVDQAKRIADL 484

RESULT 9
US-08-818-111-76
; Sequence 76, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-76

Query Match          37.2%; Score 55; DB 4; Length 580;
Best Local Similarity 40.6%; Pred. No. 2.5;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TAAADNFOLSGG--GQGFAPIGAMAIAGQI 30
Db 453 TLGADSDAQSGSIGLGFAPVDQAKRIADL 484

RESULT 10
US-09-056-556-75
; Sequence 75, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
```

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-75

Query Match 37.2%; Score 55; DB 4; Length 580;
Best Local Similarity 40.6%; Pred. NO. 2.5;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TAAASNFQLSQG--GQGAIPIGQAMAIAG 30
DB 453 TLGADSADQSSIGLGFPIPDQAKRIDEL 484

RESULT 11
US-08-895-522-3
Sequence 3, Application US/08895522
Patent No. 5858719
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
TITLE OF INVENTION: TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,522
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0336 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1167982
US-08-895-522-3

Query Match 32.1%; Score 47.5; DB 2; Length 694;
Best Local Similarity 37.1%; Pred. NO. 41;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

QY 1 TAAASN--FQLSOG-----GQGAIPIGQAMAIAG 28
DB 413 TVAFDVFHFYEYIEGQKVLNGVSFEVPAGKVAIVG 447

RESULT 12
US-09-195-391-3
Sequence 3, Application US/09195391
Patent No. 6080842
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
TITLE OF INVENTION: TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/195,391
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,522
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0336 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1167982
US-09-195-391-3

Query Match 32.1%; Score 47.5; DB 3; Length 694;
Best Local Similarity 37.1%; Pred. No. 41;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

OY 1 TAASDN--FOLSG-----GQFAIPIGQMAIAG 28
DB 413 TVAFDNVHFYEYEGQKVLGSGISFEVPAGKKVAIVG 447

RESULT 13

US-08-895-522-1
; Sequence 1, Application US/08895522
; Patent No. 5858719
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
; TITLE OF INVENTION: TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,522
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0336 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: OVARNOT02
; CLONE: 545981
; US-08-895-522-1

Query Match 31.4%; Score 46.5; DB 2; Length 747;
Best Local Similarity 37.1%; Pred. No. 64;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

OY 1 TAASDN--FOLSG-----GQFAIPIGQMAIAG 28
DB 466 TVAFDNVHFYEYEGQKVLGSGISFEVPAGKKVAIVG 500

RESULT 14

US-09-195-391-1
; Sequence 1, Application US/09195391
; Patent No. 6080842
; GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
TITLE OF INVENTION: TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/195,391
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,522
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0336 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: OVARNOT02
CLONE: 545981
US-09-195-391-1

Query Match 31.4%; Score 46.5; DB 3; Length 747;
Best Local Similarity 37.1%; Pred. No. 64;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

OY 1 TAASDN--FOLSG-----GQFAIPIGQMAIAG 28
DB 466 TVAFDNVHFYEYEGQKVLGSGISFEVPAGKKVAIVG 500

RESULT 15

US-08-846-762-90
; Sequence 90, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 344
; TYPE: PRT

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:41 ; Search time 20.7 Seconds
(without alignments)
151.037 Million cell updates/sec

Title: US-09-684-215A-18

Perfect score: 653
Sequence: 1 TAAADNFQLSGGOGFAPI.....QTKSGTGTGVTIAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCUTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	627	96.0	132	4	US-08-818-112-66 Sequence 66, Appl
2	627	96.0	132	4	US-08-818-111-67 Sequence 67, Appl
3	627	96.0	132	4	US-09-056-556-66 Sequence 66, Appl
4	627	96.0	132	4	US-08-818-112-79 Sequence 79, Appl
5	627	96.0	335	4	US-08-818-111-80 Sequence 80, Appl
6	627	96.0	335	4	US-09-056-556-79 Sequence 79, Appl
7	201.5	30.9	580	4	US-08-818-112-75 Sequence 75, Appl
8	201.5	30.9	580	4	US-08-818-111-76 Sequence 76, Appl
9	201.5	30.9	580	4	US-09-056-556-75 Sequence 75, Appl
10	158.5	24.3	97	4	US-08-818-112-72 Sequence 72, Appl
11	158.5	24.3	97	4	US-08-818-111-73 Sequence 73, Appl
12	158.5	24.3	97	4	US-09-056-556-72 Sequence 72, Appl
13	148	22.7	30	4	US-09-439-313-484 Sequence 484, App
14	128.5	19.7	460	4	US-09-199-637A-132 Sequence 132, App
15	106	16.2	475	1	US-08-278-091-6 Sequence 6, Appli
16	106	16.2	475	1	US-08-483-859-6 Sequence 6, Appli
17	106	16.2	475	1	US-08-472-173-6 Sequence 6, Appli
18	106	16.2	475	1	US-08-350-741-2 Sequence 2, Appli
19	106	16.2	475	1	US-08-487-167-6 Sequence 6, Appli
20	106	16.2	475	2	US-08-482-816-6 Sequence 6, Appli
21	106	16.2	475	2	US-08-296-149-6 Sequence 6, Appli
22	106	16.2	475	2	US-08-801-499-6 Sequence 6, Appli
23	106	16.2	475	2	US-08-463-875A-2 Sequence 6, Appli
24	106	16.2	475	2	US-08-615-271-6 Sequence 6, Appli
25	106	16.2	475	3	US-09-074-660-6 Sequence 6, Appli
26	106	16.2	475	3	US-09-074-659-6 Sequence 6, Appli
27	106	16.2	475	3	US-09-106-468-6 Sequence 6, Appli

28	106	16.2	475	4	US-09-106-466A-6 Sequence 6, Appli
29	106	16.2	475	4	US-09-106-467-6 Sequence 6, Appli
30	103.5	15.8	472	1	US-08-278-091-5 Sequence 5, Appli
31	103.5	15.8	472	1	US-08-483-859-5 Sequence 5, Appli
32	103.5	15.8	472	1	US-08-472-173-5 Sequence 5, Appli
33	103.5	15.8	472	2	US-08-482-816-5 Sequence 5, Appli
34	103.5	15.8	472	2	US-08-296-149-5 Sequence 5, Appli
35	103.5	15.8	472	2	US-08-801-499-5 Sequence 5, Appli
36	103.5	15.8	472	2	US-08-615-271-5 Sequence 5, Appli
37	103.5	15.8	472	3	US-09-074-660-5 Sequence 5, Appli
38	103.5	15.8	472	3	US-09-074-659-5 Sequence 5, Appli
39	103.5	15.8	472	3	US-09-106-468-5 Sequence 5, Appli
40	103.5	15.8	472	3	US-09-106-466A-5 Sequence 5, Appli
41	103.5	15.8	472	4	US-09-106-467-5 Sequence 5, Appli
42	103.5	15.8	472	4	US-09-106-468-5 Sequence 5, Appli
43	103.5	15.8	491	3	US-08-923-454A-14 Sequence 14, Appli
44	103.5	15.8	491	3	US-08-350-741-3 Sequence 3, Appli
45	103.5	15.8	492	2	US-08-463-875A-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-818-112-66
Sequence 66, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-66
Query Match 96.0%, Score 627, DB 4, Length 132;
Best Local Similarity 94.7%, Pred. No. 2, 4e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-6031
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-79

Query Match 96.0%; Score 627; DB 4; Length 355;
Best Local Similarity 94.7%; Pred. No. 9, 6e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

OY 1 TAAASDFOLSGGGGFAIPGAMATAGQIK---LPTVHIGTATFLGIGVNDNNGGAR 56
DB 224 TAAASDFOLSGGGGFAIPGAMATAGQIRSGGSPYHIGTATFLGIGVNDNNGGAR 283
OY 57 VORVGSAPASAGISTGDIYITAVDGAIPINSATAMADALNGHHPGDIYSVTWOTKSSGTR 116
DB 284 VORVGSAPASAGISTGDIYITAVDGAIPINSATAMADALNGHHPGDIYSVTWOTKSSGTR 343
OY 117 TGNVTLAEGPPA 128
DB 344 TGNVTLAEGPPA 355
RESULT 5
US-08-818-111-80
; Sequence 80, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-80

Query Match 96.0%; Score 627; DB 4; Length 355;
Best Local Similarity 94.7%; Pred. No. 9, 6e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

OY 1 TAAASDFOLSGGGGFAIPGAMATAGQIK---LPTVHIGTATFLGIGVNDNNGGAR 56
DB 224 TAAASDFOLSGGGGFAIPGAMATAGQIRSGGSPYHIGTATFLGIGVNDNNGGAR 283
OY 57 VORVGSAPASAGISTGDIYITAVDGAIPINSATAMADALNGHHPGDIYSVTWOTKSSGTR 116
DB 284 VORVGSAPASAGISTGDIYITAVDGAIPINSATAMADALNGHHPGDIYSVTWOTKSSGTR 343
OY 117 TGNVTLAEGPPA 128
DB 344 TGNVTLAEGPPA 355
RESULT 6
US-09-056-556-79
; Sequence 79, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392

```

; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match          96.0%; Score 627; DB 4; Length 355;
Best Local Similarity 94.7%; Pred. No. 9.6e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 TASDNFQLSGGGFAIPIGQAMATIGQK---LPTVHIGPTAFGLGVVNNNGNGAR 56
DB 224 TASNDFQLSGGGFAIPIGQAMATIGQKSGGSPPTVHIGPTAFGLGVVNNNGNGAR 283
QY 57 VORVVGSAAPASLIGSTGDTAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 116
DB 284 VORVVGSAAPASLIGSTGDTAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 343
QY 117 TGNVTLAEPPA 128
DB 344 TGNVTLAEPPA 355

RESULT 7
US-08-818-112-75
; Sequence 75, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
US-08-818-112-75

Query Match          30.9%; Score 201.5; DB 4; Length 580;
Best Local Similarity 38.0%; Pred. No. 4.2e-15;
Matches 49; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TASNDFQLSG--GGFAIPIGQAMATIGQKLPVTHIGPTAFGLGVV---DNNGNGA 55
DB 453 TLGADSDAOSGSGIGGFAIPVDQAKRIADEL---ISTGKASHASIGVQVTDKDPGA 508
QY 56 RVORVVGSAAPASLIGSTGDTAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 115
DB 509 KIVEVYAGGAANAGVPKGVVTKVDDRPINSADALVAARSKAPATVALTFQDPDSGGS 568
QY 116 RTGNVTLAE 124
DB 569 RTGVVTLGK 577

RESULT 8
US-08-818-111-76
; Sequence 76, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-76

Query Match          30.9%; Score 201.5; DB 4; Length 580;
Best Local Similarity 38.0%; Pred. No. 4.2e-15;
Matches 49; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TASNDFQLSG--GGFAIPIGQAMATIGQKLPVTHIGPTAFGLGVV---DNNGNGA 55

```

Db 453 TLGADSAQAQSGISGLGFAIPVDQAKRIADEL-----ISTGKASHASLGVQVYTNKDPTPGA 508
OY 56 RVORVGSAPASLIGISGDVITAVDGPINSATAMALNGHGGDVISWTWQKSGT 115
Db 509 KIVEVYAGGAANAGVPGVYVTKVDRPINSADALVAAVRSKAPGATVALTFQDPSGGS 568
OY 116 RTGNVTILAE 124
Db 569 RTVOVYTLGK 577

RESULT 9
US-09-056-556-75
; Sequence 75, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-75

Query Match 30.9%; Score 201.5; DB 4; Length 580;
Best Local Similarity 38.0%; Pred. No. 4,2e-15;
Matches 49; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

OY 1 TAAADNFOLSG--GGFAIPIGQAMAIAGQIKLPTVHIGPTAFLGVV--DNNNGMA 55
Db 453 TLGADSAQAQSGISGLGFAIPVDQAKRIADEL-----ISTGKASHASLGVQVYTNKDPTPGA 508
OY 56 RVORVGSAPASLIGISGDVITAVDGPINSATAMALNGHGGDVISWTWQKSGT 115
Db 509 KIVEVYAGGAANAGVPGVYVTKVDRPINSADALVAAVRSKAPGATVALTFQDPSGGS 568
OY 116 RTGNVTILAE 124
Db 569 RTVOVYTLGK 577

RESULT 10
US-08-818-112-72
; Sequence 72, Application US/08818112

; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 133
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-72

Query Match 24.3%; Score 158.5; DB 4; Length 97;
Best Local Similarity 38.7%; Pred. No. 3.3e-11;
Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

OY 35 VHTGPTAFLGLGVV--DNNNGCARVORVYGSAPASLIGISGDVITAVDGPINSATAM 91
Db 2 ISTGKASHASLGVQVYTNKDPTPGA KIVEVYAGGAANAGVPGVYVTKVDRPINSADAL 61
OY 92 ADALNGHHPGDVISTWQKSGTGTGNVTILAE 124
Db 62 VAAVRSKAPGATVALTFQDPSGGSRTVOVYTLGK 94

RESULT 11
US-08-818-111-73
; Sequence 73, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

```
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-73

Query Match      24.3%; Score 158.5; DB 4; Length 97;
Best Local Similarity 38.7%; Pred. No. 3.3e-11;
Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

QY 35 VHIGPTAFLGLGV--DNNGNGARVORYGSAPASLIGSTGDTVTAVDGAPlNSATAM 91
   : | : | | | | : | | : | | : | | : | | : | | | | | | | |
Db 2 ISTGASHASLGVQVYTNKDTPGAKIVEYVAGGAANAGVPKGVVTKVDPRINSADAL 61

QY 92 ADALNGHHPGDVISTWQTKSGTGTGNVTLAE 124
   : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 VAAVRSKAPGATVALTFDPSGSGSRVQVTLGK 94

RESULT 12
US-09-056-556-72
Sequence 72, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-72

Query Match      24.3%; Score 158.5; DB 4; Length 97;
Best Local Similarity 38.7%; Pred. No. 3.3e-11;
Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

QY 35 VHIGPTAFLGLGV--DNNGNGARVORYGSAPASLIGSTGDTVTAVDGAPlNSATAM 91
   : | : | | | | : | | : | | : | | : | | : | | | | | | | |
Db 2 ISTGASHASLGVQVYTNKDTPGAKIVEYVAGGAANAGVPKGVVTKVDPRINSADAL 61

QY 92 ADALNGHHPGDVISTWQTKSGTGTGNVTLAE 124
   : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 VAAVRSKAPGATVALTFDPSGSGSRVQVTLGK 94

RESULT 13
US-09-439-313-484
Sequence 484, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yuqun
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 484
LENGTH: 30
TYPE: PRT
ORGANISM: Homo Saplen
US-09-439-313-484

Query Match      22.7%; Score 148; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSOGGGAIFPIGQMAAINGOI 30
   : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 TAASDNFQLSOGGGAIFPIGQMAAINGOI 30

RESULT 14
US-09-199-637A-132
Sequence 132, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahne, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
```



```

; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIORITY FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIORITY FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132
; LENGTH: 460
; TYPE: PR
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-132

```

Query Match 19.7%: Score 128.5; DB 4; Length 460;

Best Local Similarity 32.4%; Pred. No. 9e-07;
Matches 44; Conservative 16; Mismatches 53; Indels 23; Gaps 4;

```

QY 4 SDNFQLSG--GGGFAIPIGQAMAIQIKLPTVHIGPTAFGLGVVDNNGN-----53
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 233 SQIFRSGGFMGLSFAPIDVALNADQLK---KAGKYSRGLGVIOEVNKKDLAESFG 288
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 54 -----GARVQRYVGSAPASLSIGTGDVITAVDGAIPNSATFAMALNHHGPDVSY-- 106
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 289 LDKPSGALVAQLEVDGDPKAPAGGLQVGLVILSLNGOSINESADLPHLVGNMKPGDKINLDV 348
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 107 --TWQTKSGGTRGNV 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 349 IRNGQRKSLSMAYGNL 364

```

```

QY 107 --TWQTKSGGTRGNV 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 349 IRNGQRKSLSMAYGNL 364

```

RESULT 15

US-08-278-091-6
Sequence 6, Application US/08278091

Patent No. 5506139

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M

APPLICANT: YANG, Yan-Ping

APPLICANT: CHONG, Pele

APPLICANT: OOMEN, Raymond P.

APPLICANT: KLEIN, Michel H.

TITLE OF INVENTION: Analog of Haemophilus Hind7 Protein with
Reduced Protease Activity

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Slim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/278,091

FILING DATE: 21-JUL-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-371

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

```

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-278-091-6

```

Query Match 16.2%: Score 106; DB 1; Length 475;

Best Local Similarity 29.6%; Pred. No. 0.00038;
Matches 37; Conservative 18; Mismatches 52; Indels 18; Gaps 3;

```

QY 13 GGGFAIPIGQAMAIQIKLPTVHIGPTAFGLGVVDNNGN-----GARVQ 59
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 264 GIGFAIPSNMVKNLTSQM---VEYGOVARGELGIMTELNSLAKAMKVDARQAFVSQ 319
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 VVGSAPASLSIGTGDVITAVDGAIPNSATFAMADALNHHGPDVSYTWQTKSGGTRTGN 119
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 320 VMPNSSAAKAGIKAGDVITSLNGKPLISFPAALRAQVGTMPVGSKISL-GLLEGRKAITVN 378
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 VYLAE 124
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 379 LELQ 383

```

Search completed: August 13, 2002, 10:35:42
Job time: 31 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:36:14 ; Search time 25.01 Seconds
(without alignments)
491.781 Million cell updates/sec

Title: US-09-684-215a-18

Perfect score: 653

Sequence: 1 TAAADNFOLSGGGGFAIRP.....OTKSGTGTGNGVTLAEGPPA 128

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.71:*

2: p1r1:*

3: p1r2:*

4: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632	96.8	355	2	F70983
2	437.5	67.0	361	2	S47170
3	422.5	64.7	354	2	A87242
4	200.5	30.7	464	2	C70821
5	185.5	28.4	382	2	H86930
6	185.5	28.4	452	2	T45448
7	140	21.4	407	2	AG2150
8	134.5	20.3	362	2	T35287
9	132.5	20.3	408	2	H86891
10	123.5	18.9	474	2	F83550
11	122.5	18.8	514	2	AB2581
12	122	18.7	394	2	S74643
13	119	18.2	441	2	E75357
14	116.5	17.8	203	2	T35866
15	115	17.6	348	2	H96956
16	112.5	17.2	530	2	F87590
17	111	17.0	499	2	B81914
18	110	16.8	455	2	C91142
19	110	16.8	455	2	F85987
20	109	16.7	393	2	E95261
21	109	16.7	397	2	B98127
22	108	16.5	429	2	AD1894
23	107	16.4	513	2	I40060
24	107	16.4	513	2	AD3418
25	106.5	16.3	459	2	F72359
26	106	16.2	455	2	AB0909
27	106	16.2	475	1	S15337
28	105.5	16.2	355	1	JC6052
29	105.5	16.2	355	2	D91142

30	105.5	16.2	355	2	G85987	protease (import
31	105.5	16.2	452	2	S77538	serine proteinase
32	105	16.1	455	2	JC6051	trypsin-like prote
33	104	15.9	475	2	AC0528	protease DO precu
34	103.5	15.8	474	2	S45229	protease DO (EC
35	103.5	15.8	474	2	B85500	protease DO (EC
36	103.5	15.8	474	2	E90649	protease DO (EC
37	103	15.8	523	2	A97479	probable serine pr
38	103	15.8	523	2	A12696	serine proteinase
39	102.5	15.7	404	2	C70886	probable integral
40	102	15.6	472	2	C87408	serine proteinase
41	102	15.6	476	2	H71936	protease DO - He
42	101	15.5	398	2	B71284	probable periplasm
43	98.5	15.1	362	2	A10433	protease (EC 3.4
44	98.5	15.1	481	2	D82826	heat shock protein
45	97.5	14.9	356	2	AC0909	serine protease (E

ALIGNMENTS

RESULT 1
F70983
probable serine proteinase pepa - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #extl_change 20-Jun-2000
C:Accession: F70983
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
R: Cole, S.T.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; MUID: 98293987
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-355 <COL>
A: Cross-references: GB:296071; GB:AL123456; NID:93242254; PIDN: CAB09453.1; PID: g21819
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: pepa
C: Superfamily: Escherichia coli trypsin-like proteinase

Query Match 96.8%; Score 632; DB 2; Length 355;
Best Local Similarity 95.3%; Pred. No. 1.3e-47;
Matches 126; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

OY 1 TAAADNFOLSGGGGFAIRP...LPTVHIGTPTAFGLGVDDNNGNGAR 56
|||||
Db 224 TAAADNFOLSGGGGFAIRP...LPTVHIGTPTAFGLGVDDNNGNGAR 283
OY 57 VQVVGSAPASLIGSTGDIYAVDGAIPNSATMADALNCHHFGDYSVTWQKSGSTR 116
|||||
Db 284 VQVVGSAPASLIGSTGDIYAVDGAIPNSATMADALNCHHFGDYSVTWQKSGSTR 343
OY 117 TGNVTLAEGPPA 128
|||||
Db 344 TGNVTLAEGPPA 355

RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #extl_change 22-Oct-1999
C:Accession: S47170
R: Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
Submitted to the EMBL Data Library, June 1993
A: Description: Isolation and characterisation of a 34Kda protein of Mycobacterium par
A: Reference number: S47170
A: Accession: S47170

Db 381 RDGSMFADVKISK 394

RESULT 10
F83550
serine proteinase MucD precursor PA0766 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83550
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: F83550
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1474 <STO>
A:Cross-references: GB:AE004511; GB:AE004091; NID:99946646; PIDN:ANG04155.1; GSPDB:GN001
C:Genetics:
A:Experimental source: strain PA01
A:Gene: mucD; PA0766

Query Match 18.9%; Score 123.5; DB 2; Length 474;
Best Local Similarity 31.6%; Pred. No. 0.0027;
Matches 43; Conservative 17; Mismatches 53; Indels 23; Gaps 4;

QY 4 SDNPLSOG--GQGFAPRIGQAMAIAGQIKLPTVHIGPTAFGLGVVDNNGN----- 53
Db 233 SQIFRSQGFEMLSFPAIPDVALNVAADQIK---KAGKYSRGVLVIEVKNDLAESFG 288
QY 54 -----GARQRYVGSAPASLSIGTGDVTAVDAPINGATAMADALNGHHPDVSY-- 106
Db 289 LDKPSGALVADLVEDGPPAKAGLQYGVILSLNGQSINSADLPHLVGNMKPKEDKINLDV 348

QY 107 --TWQKSGGRTGNV 120
Db 349 TRNGQRKSLMAVGSL 364

RESULT 11
A82581
periplasmic proteinase XF2241 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: A82581
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <STM>
A:Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAF85040.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Britones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Falcinanti, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitejima, J.P.; Klieger, J.E.; Kurama, E.E.; Laigz Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation

C:Genetics:
A:Gene: XF2241
C:Superfamily: Helicobacter serine proteinase

Query Match 18.8%; Score 122.5; DB 2; Length 514;
Best Local Similarity 36.9%; Pred. No. 0.0036;
Matches 41; Conservative 13; Mismatches 42; Indels 15; Gaps 4;

QY 4 SDNPLSOG--GQGFAPRIGQAMAIAGQIK-----LPTVHIGPTAFV---GLGVDN 50
Db 262 SQIFASGCGWGMISFPAIPNLINALNAEQIRKQGVRSMLGVEIGPIDLKKAGLGPDS 321

QY 51 NGNGARQRYVGSAPASLSIGTGDVTAVDAPINSATMADALNGHHPG 101
Db 322 --RGALVNNIPHPSPAPAKAGIEVGDIVRSNCKVISFSFDLPPLIGMPPG 370

RESULT 12
S74643
proteinase hhoA (EC 3.4.-.-) - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein s11679
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Kanevo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis.
A:Reference number: S74322; MUID:97061201
A:Accession: S74643
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:91651768; PIDN:BA16795.1; PID:9165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: hhoA
C:Superfamily: proteinase hhoB
C:Keywords: hydrolase; proteinase

Query Match 18.7%; Score 122; DB 2; Length 394;
Best Local Similarity 32.8%; Pred. No. 0.003;
Matches 43; Conservative 11; Mismatches 39; Indels 38; Gaps 5;

QY 13 GQGFAPRIGQAMAIAGQIKLPTVHIGPTAFGLGVVD-----NNGN----- 53
Db 261 GIGFAIPIDQAKAI--QNTLAAGTVPHPIYGVGMNITVDQAQNNRNPSPFIPEVD 318

QY 54 GARQRYVGSAPASLSIGTGDVTAVDAPINSATM-----ADALNGHN 99
Db 319 GILVMRYLPETPARAGIRGVDIVAVDGPDISDARLQRIEQAQINKALKLIDLKDR 378

QY 100 PGDIVSVTWQT 110
Db 379 R---LSLTWQT 386

RESULT 13
E75357
Probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75357
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J. M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaleski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896

A:Accession: E75357
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <WHI>
A:Cross-references: GB:AE002017; GB:AE000513; NID:96459527; PIDN:AAFI1312.1; PID:9645953
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1756
A:Map position: 1
C:Superfamily: proteinase hnoB

Query Match 18.2%; Score 119; DB 2; Length 441;
Best Local Similarity 28.6%; Pred. No. 0.0061;
Matches 46; Conservative 22; Mismatches 43; Indels 50; Gaps 7;

QY 6 NQIOLSGGQ-----GFAIPGQAMAIAGQIKLPTVHIGPTAFL---GLGVYDNN-----51
DB 277 NQIOLGAGAGAGAGVGFALPINTVKRL-----LPQLQAGKGGVSPSPSGVFSDLSSLP 331
QY 52 -----GNGARVQRYVGSAPASLIG-----IST-GDVITAVD 81
DB 332 PQQLKAGLPSSGALLQKYPGSPAPAAAGLRGNNKGLSPSAQTSISTDGLITAVN 391
QY 82 GAPINSATAMADALNGHHPGDVISTVQTKSGGTRGVNTL 122
DB 392 GQPLEDAGSIQEAIVLATGEGQPLRTVR--RGKTRREVETL 431

RESULT 14

T35866
probable integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T35866
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, April 1999
A:Reference number: 221591
A:Accession: T35866
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-203 <SAU>
A:Cross-references: EMBL:AL049727; PIDN:CA041567.1; GSPDB:GN00070; SCOEDB:SC9B1.21
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC9B1.21

Query Match 17.8%; Score 116.5; DB 2; Length 203;
Best Local Similarity 32.5%; Pred. No. 0.0043;
Matches 37; Conservative 15; Mismatches 49; Indels 13; Gaps 2;

QY 12 GGCGFAIPGQAMAIAGQIKLPTVHIGPT-----AFLGLGVYDNNNGARVQRY 60
DB 80 GGAGGCGATGTATGTRAPAHPSA--GPSAGRSPAPACATLGEAVEADDEKPGARVGV 137
QY 61 VGSAPASLIGTSTGVDITAVDAPINSATAMADALNGHHPGDVISTVQTKSGG 114
DB 138 HVPGPYAAGLVAGDVLAVGTTRVDSATDLAHAVARAGPEKVKLTVRHRSRG 191

RESULT 15

H96936
serine protease Do (heat-shock protein) [Imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H96936
R:Moiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H96936

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78443.1; PID:915023321; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0463

Query Match 17.6%; Score 115; DB 2; Length 348;
Best Local Similarity 28.2%; Pred. No. 0.011;
Matches 35; Conservative 25; Mismatches 48; Indels 16; Gaps 4;

QY 10 SOGGGQFAIPG-----QAMAIAGQIKLPTVHI-----GPTAFLGLGVYDNNNGARVQ 58
DB 226 SAEIGFVAIVINIVKRYLSIKTTGQFKTPVIGIIGLDSMNGVLYNL---NFEKGIIYV 281
QY 59 RVVGSAPASLIGTSTGVDITAVDAPINSATAMADALNGHHPGDVISTVQTKSGGTRTG 118
DB 282 NISPNAGAAAGINKGDIILVNGKNINININNELRESIYTTGANNVSLKIKTAS-GEKTV 340
QY 119 NVTL 122
DB 341 NVKI 344

Search completed: August 13, 2002, 10:36:15
Job time: 63 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:42:45 ; Search time 15.81 Seconds
(without alignments)
313.479 Million cell updates/sec

Title: US-09-684-215a-18

Sequence: 1 TNASDNFQUSGGGFAIRP1.....QTKSGGTRGNVTLAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.5	20.3	408	1	HTRA_LACTA
2	129	19.8	413	1	HTRA_LACH
3	114	17.5	504	1	DEGP_RHME
4	107	16.4	513	1	DEGP_BRUB
5	106	16.2	475	1	DEGP_SALT
6	105.5	16.2	355	1	DEGS_ECOLI
7	105	16.1	455	1	DEGQ_ECOLI
8	103.5	15.8	474	1	DEGP_ECOLI
9	103.5	15.8	503	1	DEGP_BARBE
10	94	14.4	478	1	DEGP_BUCAL
11	92	14.1	448	1	DEGP_ARANTH
12	88	13.5	497	1	DEGP_CHTR
13	88	13.5	905	1	ZOJ_MOUSE
14	87.5	13.4	466	1	HTRQ_HABIN
15	86	13.2	510	1	MURE_MYCTU
16	85.5	13.1	492	1	NYLA_PSES8
17	85.5	13.1	492	1	NYLA_PSES8
18	85.5	13.1	492	1	NYLA_PSES8
19	84	12.9	864	1	ELT_FAT
20	83	12.7	400	1	YTXA_BACSU
21	82	12.6	762	1	P115_CHICK
22	81	12.4	488	1	DEGP_CHIPN
23	79.5	12.2	292	1	PPNK_ECOLI
24	79.5	12.2	292	1	PPNK_ECOLI
25	79	12.0	2249	1	OMPA_RICRI
26	78.5	12.0	437	1	DEGL_ARANTH
27	78.5	12.0	543	1	7UP2_DROME
28	78.5	12.0	746	1	7UP2_DROME
29	77.5	11.9	460	1	HRA3_MOUSE
30	77.5	11.9	478	1	DEGP_BUCAP
31	77	11.8	430	1	TPSN_CHICK
32	77	11.8	2021	1	OMPA_RICN
33	76.5	11.7	219	1	VG36_BPT2

34	76.5	11.7	928	1	PMF9_CHLPP
35	75.5	11.6	475	1	TRB1_ECOLI
36	75.5	11.6	852	1	DLG2_RAT
37	75.5	11.6	1286	1	AIDA_ECOLI
38	75	11.5	467	1	ISP6_SCHPO
39	74.5	11.4	351	1	HYPE_ALCEP
40	74	11.3	314	1	YR83_MYCTU
41	74	11.3	427	1	THD1_MYCTU
42	74	11.3	434	1	CTPA_BARBA
43	73.5	11.3	452	1	HRA3_HUMAN
44	73.5	11.3	870	1	DLG2_HUMAN
45	73	11.2	711	1	SP3_HUMAN

ALIGNMENTS

RESULT	ID	STANDARD	PRT	408 AA.
1	HTRA_LACTA			
AC	O91A06			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	Serine protease do-1like htra (EC 3.4.21.-) (Htra1).			
GN	HTRA OR L2136.			
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).			
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Lactococcus			
OX	NCBI_TaxID=1360;			
RN	(1)			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RC	STRAIN-IL1403;			
RX	MEDLINE=20177820; PubMed=10712686;			
RA	Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;			
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;			
RT	"Htra is the unique surface housekeeping protease in Lactococcus			
RT	lactis and is required for natural protein processing.";			
RL	Mol. Microbiol. 35:1042-1051(2000).			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-IL1403;			
RX	MEDLINE=2125186; PubMed=11337471;			
RA	Bolotin A., Wincker P., Manger S., Jaillon O., Malarre K.;			
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;			
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus			
RT	lactis sp. lactis IL1403.";			
RL	Genome Res. 11:731-753(2001).			
CC	- FUNCTION: DEGRADABLE ABNORMAL EXPORTED PROTEINS. NEEDED FOR THE PRO-			
CC	PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF			
CC	A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED			
CC	PROTEINS.			
CC	- SUBCELLULAR LOCATION: Membrane-bound (Probable).			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE			
CC	DEGP/DEQ/DEGS FAMILY.			
CC	- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: AF155705; AA061294.1; -			
DR	EMBL: AE006442; AA06234.1; -			
DR	InterPro: IPR001478; PDZ.			
DR	InterPro: IPR001940; Protease2C.			
DR	InterPro: IPR001254; Trypsin.			
DR	Pfam: PF00595; PDZ; 1.			
DR	Pfam: PF00899; trypsin; 1.			
DR	PRINTS: PR00834; PROTEASES2C.			
DR	SMART: SM00228; PDZ; 1.			

DR PROSITE; PS50106; PDZ; 1.
 KW Hydrolyase; Serine protease; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26
 FT DOMAIN 88 284 POTENTIAL.
 FT ACT_SITE 302 383 CATALYTIC.
 FT ACT_SITE 127 127 PDZ
 FT ACT_SITE 157 157 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 239 239 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 408 AA; 41648 MW; 581B90B5A7DF851 CRC64;

Query Match
 Best Local Similarity 30.3%; Score 132.5; DB 1; Length 408;
 Matches 41; Conservative 18; Mismatches 44; Indels 31; Gaps 5;

QY 13 GGGFAIPGQMAIA-----GQIKLPVHIGPTAFGLGVVD-----NNGN----- 53
 DB 270 GLGFAIPNDVNIINKLADGKISRPA-----LGIRWVDLSQSLSTNDSSQLKIPSS 321
 QY 54 ---GARVORVGSAPASAGISTGDIYTVNDGAPINSATAMADALNGHHPGVISYVWQT 110
 DB 322 YVGGVVVYVSGSLPAPASAGLAKAGDVITKVDPAVTSSTDLQSAALYSHNINDTVVTVY- 380
 QY 111 KSGGRTGNVTAE 124
 DB 381 RDKSNTADVLSK 394

RESULT 2
 HTRA_LACHE STANDARD: PRT; 413 AA.

AC 0924H7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Serine protease do-like htra (EC 3.4.21.-).
 GN HTRA.
 OS Lactobacillus helveticus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1587;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=53/7;
 RA MEDLINE=9047559; PubMed=9829922;
 RA Smeds A., Varmann P.K., Palva A.M.;
 RT "Molecular characterization of a stress-inducible gene from
 RT Lactobacillus helveticus."
 RL J. Bacteriol. 180:6143-6153(1998).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGP/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AJ005672; CAA06668.1; -
 DR MEROPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hydrolyase; Serine protease; Transmembrane.
 FT TRANSMEM 20 40
 FT POTENTIAL.

FT DOMAIN 141 302 CATALYTIC.
 FT DOMAIN 305 401 PDZ.
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 413 AA; 42647 MW; B16B77991C8707 CRC64;

Query Match
 Best Local Similarity 31.4%; Score 129; DB 1; Length 413;
 Matches 37; Conservative 15; Mismatches 48; Indels 18; Gaps 3;

QY 8 QLSGQ-----GGFAIPGQMAIA-----GQIKLPVHIGPTAFGL-----LGVD 49
 DB 279 QSSDGTVEGMAFAIPSENEVTIVNELVKKGIKTRPOLGVYVALDGIPEGRSLIKS 338
 QY 50 NNGGARVORVGSAPASAGISTGDIYTVNDGAPINSATAMADALNGHHPGVISYV 107
 DB 339 NLRNGIYAFVSRNGSAPASAGISTGDIYTVNDGAPINSATAMADALNGHHPGVISYV 396

RESULT 3

DEGP_RHIME STANDARD: PRT; 504 AA.

ID DEGP_RHIME
 AC 052894;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 GN DEGP1 OR DEGP OR R01021 OR SMC02365.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=96146524; PubMed=8550509;
 RA Glazebrook J., Ichige A., Walker G.C.;
 RT "Genetic analysis of Rhizobium meliloti bcaA-phoA fusion results in
 RT identification of degP: two loci required for symbiosis are closely
 RT linked to degP";
 RL J. Bacteriol. 178:745-752(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelie D., Puenher A., Purnelle B., Ramsperger D.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGP/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U31512; AAC43669.1; ALT_INIT.
 DR EMBL; AL591785; CAC45593.1; -
 DR MEROPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT 172";
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
 CC SPECIFICITY WITH HHOA/DEGO.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54548; CA38420.1; -;
 DR EMBL: AE008704; AAL19173.1; -;
 DR PIR: S15337; S15337.
 DR PIR: S21327; S21327.
 DR MEROPS: S01.273; -;
 DR STYGENE: SG10173; degp.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ. 1.
 DR Pfam: PF00089; trypsin. 1.
 DR PRINTS: PRO0834; PROTEASES2C.
 DR SMART: SM00228; PDZ. 2.
 DR PROSITE: PS50106; PDZ. 2.
 DR HydroLase: Serine protease; Heat shock; Periplasmic; Signal;
 KW Complete proteome.
 FT CHAIN 1 26 BY SIMILARITY.
 FT DOMAIN 27 475 PROTEASE DO.
 FT DOMAIN 281 372 PDZ 1.
 FT DOMAIN 378 467 PDZ 2.
 FT ACT_SITE 132 132 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 475 475 AA: 49315 MW: 86E685BF3C1A289F CRC64;
 SQ SEQUENCE

Query Match 16.28; Score 106; DR 1; Length 475;
 Best local Similarity 29.68; Pred. NO. 0.066;
 Matches 37; Conservative 18; Mismatches 52; Indels 18; Gaps 3;

QY 13 GGGFAIPDIGAMAIAGIKLPTVHIGPTAFGLGVYDNNNGN-----GARVOR 59
 Db 264 GIGFAIPSNMVKNTLSQM-----VEYGVKKRGELGIMGTELSSELAKKAKVDAORAFPSQ 319
 QY 60 VYNSAPASAGISIGDVTAVDGAIPNSATPAMADALNCHHPGDVITSVTWQTKSGGTFRGN 119
 Db 320 VPMNSSAKAGIKRAGDVTISLNGKPISSFALRAOVGTMPGVSKISL-GLLRGKAIVTN 378
 QY 120 VTIAE 124
 Db 379 LELQ 383

RESULT 6
 DEGS_ECOLI
 ID DEGS_ECOLI STANDARD; PRT; 355 AA.
 AC P31137;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protease degs precursor (EC 3.4.21.-).
 CN DEGS OR HHOA OR HTRH OR B3235 OR Z4594 OR ECS4108.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Bass S., Gu Q., Goddard A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degp and degs, *Escherichia coli* genes encoding
 RT J. Bacteriol. 178:1146-1153(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., DiMantia E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 050952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genome comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
 RX MEDLINE=88105815; PubMed=332223;
 RA Vogel R.F., Ertlan K.-D., Mecke D.;
 RT "Cloning and sequence of the mdh structural gene of *Escherichia coli*
 RT coding for malate dehydrogenase.";
 RL Arch. Microbiol. 149:36-42(1987).
 RN [7]
 RP IDENTIFICATION.
 RA Bazan J.F., Fletterick R.J.;
 RT "Structural and catalytic models of trypsin-like viral proteases.";
 RL Semin. Virol. 1:311-322(1990).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U15661; AAC43992.1; -;
DR EMBL: U32495; AAC44006.1; -;
DR EMBL: U18997; AAA58037.1; -;
DR EMBL: AE000402; AAC76267.1; -;
DR EMBL: AE005551; AAG58363.1; -;
DR EMBL: AP002564; BAB37531.1; -;
DR EMBL: M24777; -; NOT_ANNOTATED_CDS.
DR MEROPS: S01.275; -;
DR Ecogene: EG11652; h0b.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ.1.
DR Pfam: PF00089; Ctrypsin.1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ.1.
DR PROSITE: PS50106; PDZ.1.
KW Hydrolase; Serine protease; Periplasmic; signal; Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 335 PROTEASE DEGS.
FT DOMAIN 281 326 PDZ.
FT ACT_SITE 96 96 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 126 126 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 253 253 R -> A (IN REF. 6).
FT CONFLICT 307 307 V -> E (IN REF. 6).
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;

Query Match 16.2%; Score 105.5; DB 1; Length 355;
Best Local Similarity 30.7%; Pred. No. 0.054;
Matches 42; Conservative 11; Mismatches 65; Indels 19; Gaps 4;

OY 6 NFOLSQGGQ-----GFAIP-----IGQAMALAGQIKLPVHIGPFAFLGLGVND 49
DB 219 SFDKNDGEPEGEGFAIPFQLAIRKIMDKLRDGRVIRIGIGREIAPLHAQGGGI-- 276
OY 50 NNGGARQQRVVGSAAPASLGISGCVITAVDGAFINATAMADLNGHPDVISVTWQ 109
DB 277 DQLGIVNEVSPDGPANMAGIQVNDLISVKNRPAISALEMTDVAIRPSVTPV--V 335
OY 110 TRSGTRTGNVLAEP 126
DB 336 MRDQKLTLYQVTFIEYF 352

RESULT 7
DEGO_ECOLI STANDARD; PRT; 455 AA.
AC P39099;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protease dego precursor (Ec 3.4.21.-).
GN DEGO OR HHOA OR B3234.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA Bass S., Gu O., Goddard A.;
RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
RC STRAIN-K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;

RA Waller P.R., Sauer R.T.;
RT "Characterization of dego and degs, Escherichia coli genes encoding
RT homologs of the DegP protease."
RL J. Bacteriol. 178:1146-1153(1996).
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9276503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1233-1238(1997).
CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U15661; AAC43992.1; -;
DR EMBL: U32495; AAC44005.1; -;
DR EMBL: U18997; AAA58036.1; -;
DR EMBL: AE000402; AAC76266.1; -;
DR MEROPS: S01.274; -;
DR SWISS-2DPAGE: P39099; COLI.
DR Ecogene: EG12612; dego.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ.2.
DR Pfam: PF00089; Trypsin.1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ.2.
DR PROSITE: PS50106; PDZ.2.
KW Hydrolase; Serine protease; Periplasmic; signal; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 455 PROTEASE DEGO.
FT DOMAIN 258 349 PDZ 1.
FT DOMAIN 355 447 PDZ 2.
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match 16.1%; Score 105; DB 1; Length 455;
Best Local Similarity 27.3%; Pred. No. 0.077;
Matches 36; Conservative 18; Mismatches 42; Indels 36; Gaps 5;

OY 13 GGGFAIPGQAMAIA-----GQIKLPVHIGPFAFLGLGVNDNNGN-----GA 55
DB 241 GIGFAIPNMARTLAQQLIDFGEIK-----RGILGKGTGEMSDIAKARNLDYQKA 292
OY 56 RQQRVVGSAAPASLGISGCVITAVDGAFINATAMADLNGHPDVISVTWQTKSGT 115
DB 293 EVSEVLPGSGSAKAGKADITISLNGKFLNSFAELRSKIATTEPOT-----KYALGL 346
OY 116 RTG-----NVTL 122
DB 347 RKGKPLEVEVTL 358

RESULT 8
DEGP_ECOLI

ID DEGP_ECOLI STANDARD: PRT; 474 AA.
 AC P09376; P15724;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE Protease do precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR PTD OR B0161 OR 20173 OR ECS0165.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=89057448; PubMed=3057437;
 RA Lipinska B., Sharma S., Georgopoulos C.;
 RT "Sequence analysis and regulation of the htra gene of Escherichia
 coli: a sigma 32-independent mechanism of heat-inducible
 transcription.";
 RL Nucleic Acids Res. 16:10053-10067(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RX MEDLINE=94261430; PubMed=8202364;
 RA Fujita N., Mori H., Yura T., Ishihama A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 2.4-4.1 min (110,917-193,643 bp) region.";
 RL Nucleic Acids Res. 22:1637-1639(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
 RA Lashari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074933; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Ten G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE OF 1-50 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=90323597; PubMed=2165018;

RA Onirk S., Bhattacharjee S.K., Bessman M.J.;
 RT "Primary structure of the deoxyguanosine triphosphate
 triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
 RL Gene 89:13-18(1990).
 RN [8]
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE=90207273; PubMed=2157212;
 RA Wurgler S.M., Richardson C.C.;
 RT "Structure and regulation of the gene for dgt triphosphohydrolase
 from Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
 RN [9]
 RP CHARACTERIZATION, AND SEQUENCE OF 27-39.
 RX MEDLINE=9020693; PubMed=2180903;
 RA Lipinska B., Zyllics M., Georgopoulos C.;
 RT "The htra (degp) protein, essential for Escherichia coli survival at
 RT high temperatures, is an endopeptidase.";
 RL J. Bacteriol. 172:1791-1797(1990).
 RN [10]
 RP IDENTITY OF HTRA AND PROTEASE DO.
 RX MEDLINE=91222240; PubMed=2025286;
 RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
 RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
 RT "Protease do is essential for survival of Escherichia coli at high
 RT temperatures: its identity with the htra gene product.";
 RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
 IC1A, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
 CC -1- SUBUNIT: MULTIMERIC.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
 TEMPERATURES ABOVE 42 DEGREES CELSIUS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M36536; AAA23994.1; -;
 DR EMBL: X12457; CAB30997.1; -;
 DR EMBL: D26562; CAB30280.1; -;
 DR EMBL: AE000125; AAC73372.1; -;
 DR EMBL: U70214; AAB08591.1; -;
 DR EMBL: AE005192; AAG54465.1; -;
 DR EMBL: AP002550; BAB33588.1; -;
 DR EMBL: M29955; AAA23717.1; -;
 DR EMBL: M31772; AAA23680.1; -;
 DR PIR: S01899; S01899.
 DR PIR: B35993; B35993.
 DR MEROPS: S01.273; -;
 DR SWISS-2DPAGE: P09376; COLI.
 DR EcoGene: E810463; degp.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 KW Hydrolase; Serine protease; Heat shock; periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 474 PROTEASE DO.
 FT DOMAIN 280 371 PDZ 1.

[illegible]

QY	110	TKSG	113
	1		
Db	430	IKRG	433
RESULT	12		
DEGP_CHLTR			
ID	DEGP_CHLTR	STANDARD;	PRT; 497 AA.
AC	P18584; O84830;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa		
DE	Immunogenic protein) (SK59).		
GN	DEGP OR HTRA OR CTR83.		
OS	Chlamydia trachomatis.		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
OX	NCBI_TaxID=813;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SEKOVAR L2;		
FX	MEDLINE=9033748; PubMed=2379836;		
RA	Kahane S., Weinstein Y., Sarov I.;		
RT	"Cloning, characterization and sequence of a novel 59-kDa protein of		
RT	Chlamydia trachomatis.";		
RL	Gene 90:61-67(1990).		
LN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=D/UM-3/CX;		
FX	MEDLINE=9900809; PubMed=9784136.		
RA	Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,		
RA	Mitchell W.P., Olinger L., Ratunov R.L., Zhao Q., Koonin E.V.;		
RT	Davis R.W.;		
RT	"Genome sequence of an obligate intracellular pathogen of humans:		
RT	Chlamydia trachomatis.";		
RT	Science 282:754-759(1998).		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY SZC; ALSO KNOWN AS THE		
CC	DEGP/DEGO/DEGS FAMILY.		
CC	-1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.		
CC	-1- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS		
CC	THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN		
CC	ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL		
CC	PROTEIN.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its contact is in no way		

Tue Aug 13 10:27:59 2002

us-09-684-215a-18.rsp

Page 11

Search completed: August 13, 2002, 10:42:46
Job time: 389 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:42:23 ; Search time 43.59 Seconds
(Without alignments)
507.992 Million cell updates/sec

Title: US-09-684-215a-18
Perfect score: 653
Sequence: 1 TRASNPFOLSGGCGFAIP1.....QTRSGGTRGNVTLAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mmc:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632	96.8	355	16	007175 mycobacteri
2	437.5	67.0	361	2	050320 mycobacteri
3	422.5	64.7	354	16	09C9Y mycobacteri
4	200.5	30.7	464	16	053896 mycobacteri
5	185.5	28.4	382	16	09CDB7 mycobacteri
6	185.5	28.4	452	2	092566 mycobacteri
7	163.5	25.0	542	2	09F8K9 streptomyce
8	149	22.8	519	2	093J30 streptomyce
9	134.5	20.6	474	2	09S2R5 streptomyce
10	128.5	19.7	474	2	09ALSI pseudomonas
11	123.5	18.9	371	2	031388 bradyrhizob
12	123.5	18.9	474	16	057155 pseudomonas
13	122.5	18.8	514	16	09PBA3 xyella fas
14	122	18.7	394	16	P72780 synechoyct
15	120	18.4	473	2	044476 azotobacter
16	119	18.2	441	16	Q9RTK4 delinococcus

17	116.5	17.8	203	2	Q9X840 streptomyce
18	115	17.6	348	16	Q97L01 clostridium
19	114	17.5	513	16	Q98CS8 rhizobium l
20	112.5	17.2	530	16	Q9A4S2 caulobacter
21	111	17.0	499	16	Q9JVT1 neisseria m
22	110	16.8	481	2	Q9A0D1 pseudomonas
23	110	16.8	516	16	Q985F9 rhizobium l
24	109.5	16.8	465	16	Q920E6 rhizobium m
25	109	16.7	393	16	Q97N37 streptococc
26	109	16.7	397	2	Q06670 streptococc
27	106.5	16.3	459	16	Q9W241 thermotoga
28	106	16.2	500	2	Q44652 bruceella ab
29	105.5	16.2	452	16	P73354 synechoyct
30	105.5	16.2	491	2	Q9L8K0 shigella so
31	103.5	15.8	407	16	Q99XG9 streptococc
32	103	15.8	397	2	Q93F87 streptococc
33	102.5	15.7	404	16	Q33351 mycobacteri
34	102	15.6	472	16	Q9A8R9 caulobacter
35	102	15.6	476	16	Q92M18 helicobacte
36	101	15.5	398	16	Q83752 treponema p
37	100.5	15.4	453	2	Q9FD11 aeromonas h
38	100	15.3	500	2	Q9KJN6 myxococcus
39	98.5	15.1	481	16	Q9PGJ3 xyella fas
40	97.5	14.9	428	16	Q98N31 rhizobium l
41	97.5	14.9	474	2	Q44596 bacillus su
42	96.5	14.8	224	16	Q35021 bacillus su
43	96.5	14.8	458	2	Q9R911 bacillus su
44	95.5	14.6	433	16	Q97GD5 clostridium
45	95.5	14.6	453	16	Q67436 aquifex aeo

ALIGNMENTS

RESULT 1
007175 PRELIMINARY; PRT; 355 AA.
AC 007175;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 34.9 KDA PROTEIN.
GN PEPA OR RV0125 OR MFC1A18B.07.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID-1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE-98295987; PubMed-9634230; Garnier T., Churcher C., Harris D., Cole S.T., Brosch R., Parkhill J., Barry C.E. III, Tekala F., Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares K., Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RT Nature 393:537-544(1998).
RL -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL; Z96071; CAB09453.1; -
DR MEROPS; S01.00C; -
DR Tuberculin; RV0125; -
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PR00595; PDZ; 1.
DR Pfam; PR00089; Trypsin; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.

DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN.1.
 KW Complete proteome; Hydrolase; Hypothetical protein; Serine protease.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.8%; Score 632; DB 16; Length 355;
 Best Local Similarity 95.5%; Pred. No. 3.6e-41;
 Matches 126; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

OY 1 TAASDNFOLSGGGGFALPIGQAMAIAGQIK----LPTVHIGPTAFGLGVYDNNNGGAR 56
 DB 224 TAASDNFOLSGGGGFALPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVYDNNNGGAR 263
 OY 57 VQRVVGSAPAAASLGISTGDIYTAVDGAPINSATAMADALNGHHPGDIYSVTWQTKSGGTR 116
 DB 284 VQRVVGSAPAAASLGISTGDIYTAVDGAPINSATAMADALNGHHPGDIYSVTWQTKSGGTR 343
 OY 117 TGNVTLAEGPPA 128
 DB 344 TGNVTLAEGPPA 355

RESULT 2
 ID Q50320 PRELIMINARY; PRT; 361 AA.
 AC Q50320;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 34KDA PROTEIN PRECURSOR.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JDB8/107;
 RX MEDLINE=95005449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterisation of a putative serine protease
 expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
 RL Microbiology 140:1977-1982(1994).
 DR EMBL: Z230932; CAA80638.1; -;
 DR MEROPS; S01.0PC; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM0028; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR Hydrolase; Serine protease; Signal.
 KW SIGNAL 1
 FT SIGNAL 38
 FT SIGNAL 38
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 67.0%; Score 437.5; DB 2; Length 361;
 Best Local Similarity 65.9%; Pred. No. 3.4e-26;
 Matches 87; Conservative 18; Mismatches 22; Indels 5; Gaps 2;

OY 1 TAASDNFOLSGGGGFALPIGQAMAIAGQIK----LPTVHIGPTAFGLGVYDNNNGGAR 56
 DB 231 TAATDSTKMS-GGGGFALPIGRAMAVANQIRSGAGSNTVHIGPTAFGLGVYDNNNGGAR 289
 OY 57 VQRVVGSAPAAASLGISTGDIYTAVDGAPINSATAMADALNGHHPGDIYSVTWQTKSGGTR 116
 DB 290 VQRVVNTGPAAMAGIAGDIYTGDIYVINGATSTEVLYVPHHPGDIYVHFRSYDGGGR 349
 OY 117 TGNVTLAEGPPA 128
 DB 350 TANITTLAEGPPA 361

RESULT 3
 ID Q9CCY9 PRELIMINARY; PRT; 354 AA.
 AC Q9CCY9;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PROBABLE SECRETED SERINE PROTEASE.
 GN M2659.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1769;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Bigmiller K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felkell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AL583926; CAC32191.1; -;
 DR MEROPS; S01.0PC; -;
 DR Leproma; ML2659; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR00126; Ser-Proteas_V8.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR PRINTS; PR00839; VBPROTEASE.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EAAA CRC64;

Query Match 64.7%; Score 422.5; DB 16; Length 354;
 Best Local Similarity 62.9%; Pred. No. 4.7e-25;
 Matches 83; Conservative 18; Mismatches 26; Indels 5; Gaps 2;

OY 1 TAASDNFOLSGGGGFALPIGQAMAIAGQIK----LPTVHIGPTAFGLGVYDNNNGGAR 56
 DB 224 TAATDNTKMK-LGQGFALPIGQAMEVVGAIIRSGAGSNTVHIGPTAFGLGVYDNNNGGAR 282
 OY 57 VQRVVGSAPAAASLGISTGDIYTAVDGAPINSATAMADALNGHHPGDIYSVTWQTKSGGTR 116
 DB 283 VAVRVNTGPAAMAGIAGDIYTGDIYVISEATAMTVLVVPHHNGDIYVAVVYRSGAGGDL 342
 OY 117 TGNVTLAEGPPA 128
 DB 343 TANVTLAEGPPA 354

RESULT 4
 ID Q53896 PRELIMINARY; PRT; 464 AA.
 AC Q53896;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE SERINE PROTEASE.
 GN RV0983 OR MT0044.11.

OC Mycobacteria tuberculosis.
OC Bacteria, Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,
RA Badoec K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Honysky T., Jajels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrett L.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -! SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AL021939; CAA17362.1; -.
DR MEROPS: S01.0PC; -.
DR TubercuList; Rv0983; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS0106; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease, Serine protease.
KW SEQUENCE 464 AA; 46452 MW; A693BFC53EEEC8F CRC64.

Query Match	30.7%	Score 200.5	DB 16	Length 464
Best Local Similarity	38.0%	Pred. No. 7.8e-08		
Matches 49	Conservative 21	Mismatches 50	Indels 9	Gaps 3
OY	1	TAASNFQLSOC--GQGAIRPIGAMAIAGQIKLPTVHIGIPAFGLGVV--DNNNGCA	55	
		: : : : : : : : : : : :		
Db	337	TLGASADAOSSISLIGRAIPYDQAKRIADL-----ISGKASHSLGQVYINDKDTLGA	392	
OY	56	RVORVGSNAPASALSIGTSGDVITAAVDGAPISATAMADALNCHHPGDVTSYWTQKSGCT	115	
		:: : : : : : : : : : : : :		
Db	393	KIVEVVGAGAAANACVPKGVVYTKYDDRRINSADLVAAVRSKAPGATVAFLFQDPSGGS	452	
OY	116	RTGNVTLAE 124		
		:		
Db	453	RTVQVTLGK 461		
RESULT	5			
O9CD67				
ID	O9CD67	PRELIMINARY:	PRT:	382 AA.
AC	O9CD67:			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	POSSIBLE SECRETED SERINE PROTEASE.			
GN	ML0176.			
OS	Mycobacterium leprae.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1769;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP				
RC	STRAIN=TN;			
RX	MEDLINE=21128732; PubMed=11234002;			
RA	Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,			
RA	Wheeler P.R., Honore N., Garnier T., Churche C., Harris D.,			
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,			
RA	Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,			

RA Murphy L., Oliver K., Quall M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrett B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 403:1007-1011(2001)
CC -1 STIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583917; CAC29684.1; -.
DR MEROPS; S01.0PC; -.
DR Leptoma; ML0176; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; SM00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; hydrolase; Protease; Serine protease.
SQ SEQUENCE 362 AA; 37084 MW; 3DD8DDDAE32A80D CRC64;

```

Query Match Similarity      28.4% Score 185.5; DB 16; Length 382;
Best Local Alignment      38.1% Pred. No. 8.9e-07;
Matches 48; Conservative 18; Mismatches 53; Indels 7; Gaps

OY      2 AASDFQLSGSGGCGEFAIRPGQAMAIAGQIKLPTVHIGPTAFLEGV--VDNNGNGARVQ 58
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      258 ADSGDAQSGSGIGLGFALRIVDQAKRIADLEL---ISTGKATHASLGVOVADPKGTPGAKM 313
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      59 RYVGSAAPAASLIGISTGDIYITAVDCAPINSATAMADALNGHHPRGVISVTYQTKSGGRTG 118
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      314 DVVAGGAANAANAAPVPGVGLTRKVDRLISSADALVAAYKSAKPGKSVSLTYODSGSSRTV 373
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      119 NVTLAE 124
      | | | |
DB      374 QVTLKK 379

RESULT      6
0925G6      PRELIMINARY;      PRT;      452 AA.
AC      0925G6;
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      PUTATIVE SERINE PROTEASE.
GN      MCB373.28.
OS      Mycobacterium leprae.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RX      NCBI_TaxID=1769;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Harris D., Taylor K.;
RL      Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL      Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93188700; PubMed=8446027;
RA      Elgmelzer K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT      "Use of an ordered cosmid library to deduce the genomic organization
      of Mycobacterium leprae.";
RL      MOL. MICROBIOL. 7:197-206(1993).
DR      EMBL; AL035500; CAB36690.1; -.
DR      MEROPS; S01.UPC; -.
DR      InterPro; IPR001478; PDZ.
DR      InterPro; IPR001940; Protease2C.
DR      InterPro; IPR001254; Trypsin.
DR      Pfam; PF00595; PDZ; 1.

```


ID	Q952R5	PRELIMINARY:	PR1:	362 AA.
AC	Q952R5;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	PURATIVE SECRETED PROTEASE.			
GN	SC5F7.30.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID:1902;			
NP	11			
NP	SEQUENCE FROM N.A.			
RC	STRAIN-A3(2);			
RC	Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
RN	13			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-A3(2);			
RX	MEDLINE:97000351; PubMed=8843436;			
RA	Redenbach M., Kieser H.M., Denaparte D., Eichner A., Cullum J.,			
RA	Klinasht H., Hopwood D.A.;			
RT	"A set of ordered cosmids and a detailed genetic and physical map for			
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";			
RL	Mol. Microbiol. 21:77-86(1996).			
DR	EMBL; AL096872; CAB51255.1; -.			
DR	MEROPS; S01.0PC; -.			
DR	InterPro; IPR001478; PDZ.			
DR	InterPro; IPR001940; Protease2C.			
DR	InterPro; IPR001254; Trypsin.			
DR	Pfam; PF00595; PDZ. 1.			
DR	Pfam; PF00089; Trypsin. 1.			
DR	PRINTS; PR00834; PROTEASE2C.			
DR	SMART; SM00228; PDZ. 1.			
DR	PROSITE; PS50106; PDZ. 1.			
DR	PROSITE; PS50240; TRYPsin.DOM. 1.			
KW	Hydrolase; Protease; Serine protease			
SO	SEQUENCE 362 AA; 36381 MW; 71CC41F943ABD03B CRC64;			
Query Match 20.6%; Score 134.5; DB 2; Length 362;				
Best Local Similarity 35.2%; Pred. NO. 0.007;				
Matches 43; Conservative 14; Mismatches 50; Indels 15; Gaps 5;				
QY	13 GGGFAPIPGQAMATGQIKLPVHIGPT-----ATLGL--GVVDNNGN--GARVQRVVG 62			
DB	245 GIGFAPIPASMTVTVAQGI---VRDGKVTDSGRAALGIARTAVDVSDYRPAAGAAVEVSD 300			
QY	63 SAPAFSLISGSDGTTVAADGAPINATAMADALNGHHPGDVSVTWOTSGSGRTGNVL 122			
DB	301 GGAADDAQLRQDVLVKLCGDITITITISLEALASMRPRDKIKVY-TRDGEHETAEVTL 359			
QY	123 AE 124			
DB	360 GE 361			
RESULT 10				
Q9ALSI	PRELIMINARY:	PRT;	474 AA.	
AC	Q9ALSI;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	MUCD.			
GN	MUCD.			
OS	Pseudomonas aeruginosa.			

Query Match	Best Local Similarity	Score	DB	Length
Matches 44; Conservative 16; Mismatches 53; Indels 23; Gaps	19.7%; 32.4%;	128.5; 0.028;	DB 2;	474;
4 SDNFDLSOG--GCGAFAIIGOMAIAGQIKLPVHIGTPTAFGLGVVDNNGN-----	53			
233 SGIFFRRSGGFMCSTAFRTIDVALNADQLK-----KAGVNSGWLGVYIQEYNNDLASFQ	288			
54 ----GAAVQVRVSGAPASISGISTGDVITYAVDCAPIINSATAMADALNGHPGVISV--	106			
289 LDKPFGALVAOLVDEGPAKAGLGVGYDILSNGSINESADLPFLVGNMKPGDKINLDV	348			
107 --TWOTSGSGTGTGNY 120				
349 IRNGQRKSLSMVAGNL 364				
RESULT 11				
031388 PRELIMINARY; PRT; 371 AA.				
AC 031388;				
DF 01-JAN-1998 (TREMBLrel. 05, Created)				
DF 01-JAN-1998 (TREMBLrel. 05, Last sequence update)				
DF 01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE DEGP PROTEIN.				
GN DEGP.				
OS Bradyrhizobium japonicum.				
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC Bradyrhizobium group; Bradyrhizobium.				
OX NCBI_TaxID=375;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=98114461; PubMed=9446679;				
RA Natterhaus F., Weighofer W., Fischer H.M., Hennecke H.				
RT "Identification of the Bradyrhizobium japonicum degp gene as part of				
RT an Operon containing small heat shock protein genes.";				
RL Arch. Microbiol. 169:89-97(1998).				
DR EMBL; AF343973; AAK11276.1; -				
DR InterPro; IPR001478; PDZ.				
DR InterPro; IPR001254; Trypsin.				
DR Pfam; PF00595; PDZ; 2.				
DR PRINTS; PR00834; PROTEASES2C.				
DR SMART; SM00228; PDZ; 2.				
DR PROSITE; PS50240; TRYPsin_DOM; 1.				
KM Hydrolase; Serine protease.				
SO SEQUENCE 474 AA; 50349 MW; A22FD4338B859DAC CRC64;				

SQ SEQUENCE 371 AA; 39108 MW; 3CE1C77E17B86CDC CRC64;
 Query Match 18.9%; Score 123.5; DB 2; Length 371;
 Best Local Similarity 26.6%; Pred. No. 0.051;
 Matches 33; Conservative 26; Mismatches 48; Indels 17; Gaps 5;
 OY 13 GGGFAIPIGQAMAIAGQIKLPVHIGPT--AFILGLGYVD-----NNGNGARQVRVYG 62
 DB 243 GIGFAVPIIMARVMDL-----VOYGEVRGQIGISIRDLGVDLAKESVGLALIAELAS 298
 OY 63 SAPAASLGSTGVITAVDCAPIINSATAMADALNGHHPGVISTWQTKSGGTGTGAVTL 122
 DB 299 GSPAEGAGLQKGDIVAKVADGTPIRASQLENLILGLTPVGSRVLELRF--RNGAARSASVYEV 357
 OY 123 AEGP 126
 DB 358 --CP 359
 RESULT 12
 057155 PRELIMINARY; PRT; 474 AA.
 AC 057155;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MUCD (SERINE PROTEASE MUCD).
 GN MUCD OR PA0766.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=93391358; PubMed=8378309;
 RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,
 RA Deretic V.;
 RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa
 RT infecting cystic fibrosis patients.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=95286510; PubMed=7768826;
 RA Yu H., Schurr M.J., Deretic V.;
 RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
 RT aeruginosa AlgU: E. coli RpoE restores mucoidy and reduces sensitivity
 RT to reactive oxygen intermediates in algU mutants of P. aeruginosa.";
 RL J. Bacteriol. 177:3259-3268(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=96134987; PubMed=8550474;
 RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,
 RA Deretic V.;
 RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas
 RT aeruginosa in cystic fibrosis encode homologs of the serine protease
 RT HtrA.";
 RL J. Bacteriol. 178:511-523(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stoyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Tian Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lapidg K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";

RL Nature 406:959-964(2000).
 CC -I SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; U49151; AAC43718.1; -.
 DR EMBL; U32853; AAC43676.1; -.
 DR EMBL; AE004511; AAG04155.1; -.
 DR MEROPS; S01.UPC; -.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 DR PROSITE; PS50240; TRYPSIN DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 474 AA; 50321 MW; 8AB5D2A8967BEBE CRC64;
 Query Match 18.9%; Score 123.5; DB 16; Length 474;
 Best Local Similarity 31.6%; Pred. No. 0.067;
 Matches 43; Conservative 17; Mismatches 53; Indels 23; Gaps 4;
 OY 4 SDNFQLSG--GGGFAIPIGQAMAIAGQIKLPVHIGPTAFILGLGYVDNNGN----- 53
 DB 233 SQIFPRSGFMGLSFAIPIDVALNVADOLK---KAGKVSRLGIVQIEVNKDLAESFG 288
 OY 54 -----GARVORYGASAPASLGSTGVITAVDCAPIINSATAMADALNGHHPGVISTV-- 106
 DB 289 LDKPSGALVAQLVEDGPPAKGLQYGVYLLSLNGSINSDILPHLVGNMKKGTINLIV 348
 OY 107 --TWQTKSGGTGTGNV 120
 DB 349 IRNGGRKSLMAVGS 364
 RESULT 13
 09PBA3 PRELIMINARY; PRT; 514 AA.
 ID 09PBA3;
 AC 09PBA3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PERIPLASMIC PROTEASE.
 GN Xf2241.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colaço N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Falcinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Gandler M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miranda E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M.Y., Meidanis J., Setubal J.C.;
 RT "the genome sequence of the plant pathogen *Xylella fastidiosa*.";
 RU Nature 406:151-159(2000).
 CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AE004037; AAF85040.1; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR00126; Ser_proteas_V8.
 DR InterPro: IPR001254; trypsin.
 DR Pfam: PF00395; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PRINTS: PR00839; V8PROTEASE.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Serine protease.
 SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BE4C CRC64;

Query Match 18.8%; Score 122.5; DB 16; Length 514;
 Best Local Similarity 36.9%; Pred. No. 0.088;
 Matches 41; Conservative 13; Mismatches 42; Indels 15; Gaps 4;

QY 4 SDNFOLSG--GQGFALPIGQAMAIAGQIK-----LPVHIGPTAFL--GIGVVDN 50
 DB 262 SQIFSGAGVGWISFALPIFALINAEQIRTKGVQRBSMLGVEIFIDALKKGGIGLPPDS 321
 QY 51 NGGNARQVRVGSAPASLSIGTSDVITAVDCAPINSATAMADALNGHHHPG 101
 DB 322 --REALVNNIPRHPSPAKAGIEVDVIRSVNGKVISFSFDLPILIGMPPG 370

RESULT 14
 P72780 PRELIMINARY; PRT; 394 AA.
 AC 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE PROTEASE HHOA.
 GN HHOA OR SLI1679.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Matanabe A., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: D90900; BAA16795.1; -.
 DR MEROPS: S01.274; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PSS0106; PDZ; 1.
 DR PROSITE: PSS0106; PDZ; 1.

DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 394 AA; 41336 MW; AD59D94811B8F57B CRC64;

Query Match 18.7%; Score 122; DB 16; Length 394;
 Best Local Similarity 32.8%; Pred. No. 0.071;
 Matches 43; Conservative 11; Mismatches 39; Indels 38; Gaps 5;

QY 13 GQGFALPIGQAMAIAGQIKLPVHIGPTAFLGIGVVD-----NGN----- 53
 DB 261 GIGFALPIDAKAI--ONTLAGTVPHPYIGVQMMNTTVDQAQNNRNPSPFIPEVD 318
 QY 54 GARQVRVGSAPASLSIGTSDVITAVDCAPINSATAM-----ADALNGHH 99
 DB 319 GILVMRVLPGRPAERAGIRKRDVIAVDGTPISDGARLQRIVEAGLKNALKLIDRGDR 378
 QY 100 PGDIVSVTWQT 110
 DB 379 R--LSLTVO 386

RESULT 15
 Q44476 PRELIMINARY; PRT; 473 AA.
 AC Q44476;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MUCD.
 OS Azotobacter vinelandii.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Azotobacter.
 OX NCBI_TaxID=354;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UW 136;
 RX MEDLINE=96178940; PubMed=8606151;
 RA Martinez-Salazar J.M., Moreno S., Najera R., Boucher J.C., Espin G.,
 RA Soberon-Chavez G., Deretic V.;
 RT "Characterization of the genes coding for the putative sigma factor
 RT Algu and its regulators MucA, MucB, MucC, and MucD in Azotobacter
 RT vinelandii and evaluation of their roles in alginate biosynthesis.";
 RL J. Bacteriol. 178:1800-1808(1996).
 DR EMBL: U30799; AAB01513.1; -.
 DR MEROPS: S01.UPC; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 473 AA; 50282 MW; C97B357D897738AB CRC64;

Query Match 18.4%; Score 120; DB 2; Length 473;
 Best Local Similarity 31.2%; Pred. No. 0.12;
 Matches 43; Conservative 16; Mismatches 59; Indels 20; Gaps 4;

QY 4 SDNFOLSG--GQGFALPIGQAMAIAGQIKLPVHIGPTAFLGIGVVDNNGN----- 53
 DB 233 SQIFRSGGFMGISFALPIEVAMGVADOLKA---TGKVARGMVIGVIOEVNKNDLAESFG 288
 QY 54 -----GARQVRVGSAPASLSIGTSDVITAVDCAPINSATAMADALNGHHHPGDIVSVTW 108
 DB 289 LDRPAGALVAQVLEDSRPAKGGIQAQVDVILSDGNPIVMSADLPHLVGGLKPGAAANLE- 347
 QY 109 QTKSGTRTGNVTLAEGP 126
 DB 348 VVRDGKRRNIATVGAIP 365

Tue Aug 13 10:28:00 2002

us-09-684-215a-18.rpt

Page 8

Search completed: August 13, 2002, 10:42:25
Job time: 398 sec
